BIOMETRIKA

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF
BIOLOGICAL PROBLEMS

FOUNDED BY
W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY
EGON S. PEARSON

IN CONSULTATION WITH

HARALD CRAMER	J. B. S. HALDANE
R. C. GEARY	G. M. MORANT
MAJOR GREENWOOD	JOHN WISHART

VOLUME XXXI

1939

ISSUED BY THE BIOMETRIKA OFFICE
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

First printed in Great Britain at the University Press, Cambridge
Reprinted by offset-litho by Percy Lund Humphries & Co. Ltd

CONTENTS OF VOLUME XXXI

Memoirs

		PAGES
I.	Random and Systematic Arrangements. By HAROLD JEFFREYS	1–8
II.	A Note on Normal Correlation. By E. J. G. PITMAN	9–12
III.	A Test for the Significance of the Difference between the two Variances in a Sample from a Normal Bivariate Population. By W. A. MORGAN. With one Figure in the Text	13–19
IV.	The Distribution of Range in Samples from a Normal Population, expressed in Terms of an Independent Estimate of Standard Deviation. By D. NEWMAN. With one Figure in the Text	20–30
V.	On a Comprehensive Test for the Homogeneity of Variances and Covariances in Multivariate Problems. By D. J. BISHOP	31–55
VI.	Observed and Theoretical Ratios in Mendelian Inheritance. By E. ROBERTS, W. M. DAWSON and MARGARET MADDEN	56–66
VII.	Note on the Preceding Analysis of Mendelian Segregations. By J. B. S. HALDANE	67–71
VIII.	The Use of Statistical Methods in the Investigation of Problems of Classification in Anthropology. Part I. The General Nature of the Material and the Form of Intra-racial Distributions of Metrical Characters. By G. M. MORANT. With three Figures in the Text	72–98
IX.	A Study of the Cranial and other Human Remains from Palestine Excavated at Tell Duweir (Lachish) by the Wellcome-Marston Archaeological Research Expedition. By D. L. RISDON. With ten Figures in the Text, twenty-three Plates and six tissues in cover pocket	99–166
X.	Estimating Bacterial Populations by the Dilution Method. By ROBERT DEAN GORDON	167–180

PAGES

XI. Note on the Inverse and Direct Methods of Estimation in R. D. Gordon's problem. By E. S. PEARSON . . . 181–186

XII. On the Distribution of Maximum Likelihood Estimates. By B. L. WELCH 187–190

XIII. On Neyman's "Smooth" Test for Goodness of Fit. I. Distribution of the Criterion ψ^2 when the Hypothesis Tested is True. By F. N. DAVID. With one Figure in the Text 191–199

XIV. Tests of Hypotheses Concerning Location and Scale Parameters. By E. J. G. PITMAN 200–215

XV. On Generalized Analysis of Variance. (I.) By P. L. HSU 221–237

XVI. The Derivation of the Fifth and Sixth Moments of the Distribution of b_2 in Samples from a Normal Population. By C. T. HSU and D. N. LAWLEY. With nine Figures in the Text 238–248

XVII. Testing the Homogeneity of a Set of Variances. By H. O. HARTLEY 249–255

XVIII. The Simultaneous Distribution in Samples of Mean and Standard Deviation, and of Mean and Variance. By L. TRUKSA. With seven Figures in the Text 256–271

XIX. Certain Projective Depth and Breadth Measurements of the Facial Skeleton in Man. By ALETTE SCHREINER. With two Figures in the Text 272–286

XX. Transposition of the Viscera and other Reversals of Symmetry in Monozygotic Twins. By E. A. COCKAYNE. With two Figures in the Text and two Plates . . 287–294

XXI. The Human Remains of the Iron Age and Other Periods from Maiden Castle, Dorset. By C. N. GOODMAN and G. M. MORANT. With three Plates 295–312

XXII. Homogeneity of Results in Testing Samples from Poisson Series. With an Application to Testing Clover Seed for Dodder. By J. PRZYBOROWSKI and H. WILENSKI. With three Figures in the Text 313–323

XXIII. On the Method of Paired Comparisons. By M. G. KENDALL and B. BABINGTON SMITH. With one Figure in the Text 324–345

XXIV. The Mean and Variance of χ^2, when used as a Test of Homogeneity, when Expectations are Small. By J. B. S. HALDANE 346–355

		PAGES
XXV.	A Note on the Statistical Analysis of Sentence-length as a Criterion of Literary Style. By C. B. WILLIAMS. With five Figures in the Text	356–361
XXVI.	Applications of the Non-central t-distribution. By N. L. JOHNSON and B. L. WELCH	362–389

Miscellanea

(i)	On the Calculation of the Cumulants of the χ-distribution. By N. L. JOHNSON and B. L. WELCH	216–218
(ii)	Note on Discriminant Functions. By B. L. WELCH	218–220
(iii)	Review of E. W. Sinnott and L. C. Dunn's *Principles of Genetics*	220
(iv)	Corrections to Formulae in Papers on the Moments of χ^2. By J. B. S. HALDANE	220
(v)	Stirling's Formula with Remainder. By E. V. HUNTINGTON	390
(vi)	A Note on the Interpretation of Quasi-sufficiency. By M. S. BARTLETT	391–392
(vii)	The Cumulants and Moments of the Binomial Distribution, and the Cumulants of χ^2 for a $(n \times 2)$-fold Table. By J. B. S. HALDANE	392–396
(viii)	Review of A. A. Tschuprow's *The Principles of the Mathematical Theory of Correlation*	396

List of Plates and Folding Tables

D. L. Risdon: A Study of the Cranial and Other Human Remains from Palestine Excavated at Tell Duweir (Lachish) by the Wellcome-Marston Archaeological Research Expedition.

Ten folding tables of individual measurements of Lachish Skulls		*to face page* 162
Plate I.	Views of the Interior of Tomb 120	„ 166
Plate II.	Tomb 120 showing Collections of Skulls	„ 166
Plate III.	Exceptional Crania	„ 166

Plate IV. A Cranium showing Injuries received during Life *to face page* 166
Plate V. Two Trepanned Crania . . . . ,, 166
Plate VI. Two Artificially Deformed Crania . . . . ,, 166
Plate VII. Six Deformed Crania . . . . ,, 166
Plate VIII. Three Anomalous Crania . . . . . . ,, 166
Plate IX. Two Typical Crania: *norma lateralis* . . . ,, 166
Plate X. Two Typical Crania: *norma facialis* . . . ,, 166
Plate XI. Two Typical Crania: *norma verticalis* . . . ,, 166
Plate XII. Two Typical Crania: *norma occipitalis* . . ,, 166
Plate XIII. Five Crania with Sagittal Suture Prematurely Obliterated . . . . ,, 166
Plate XIV. Three Male Crania with Sutural Anomalies . ,, 166
Plate XV. Anomalous Regions of Two Crania . . . ,, 166
Plate XVI. A Cranium of Unusual Form . . . . . ,, 166
Plate XVII. A Diseased Cranium . . . . ,, 166
Plate XVIII. Four Exceptional Crania . . . . . . ,, 166
Plate XIX. A Molar Tooth with a Filling . . . . . ,, 166
Plate XX. Anomalous Jaws of a Skull . . . . . ,, 166
Plate XXI. Palates with Anomalous Dentitions . . . . ,, 166
Plate XXII. Jaws with Anomalous Dentitions . . . . ,, 166
Plate XXIII. Jaws with Anomalous Dentitions or Cysts . . ,, 166
Six tissues in cover pocket

E. A. Cockayne: Transposition of the Viscera and Other Reversals of Symmetry in Monozygotic Twins.

Plate I. Left and Right Profiles of Eileen and Joan . . . ,, 294
Plate II. Full-face Portraits of Eileen and Joan . . . ,, 294

C. N. Goodman and G. M. Morant: The Human Remains of the Iron Age and other Periods from Maiden Castle, Dorset.

Two folding tables of individual measurements of crania and long bones from Maiden Castle . . . . ,, 312
Plate I. Three Aspects of a Typical Male Skull from the Belgic War Cemetery . . . . ,, 312
Plate II. Exceptional Skulls from the Belgic War Cemetery . ,, 312
Plate III. Skulls from Maiden Castle with Dental and Other Anomalies . . . . ,, 312

Vol. XXXI. Parts I and II July, 1939

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY

EGON S. PEARSON

IN CONSULTATION WITH

Reprinted by offset-litho, [illegible]

ISSUED BY THE BIOMETRIKA OFFICE
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

PRINTED IN GREAT BRITAIN

[*Issued* 25 *July* 1939]

RANDOM AND SYSTEMATIC ARRANGEMENTS

By HAROLD JEFFREYS, F.R.S.

I HAVE followed with interest the recent discussions of this question, though I cannot claim personal knowledge of the types of problem where the proposed arrangements are used. My personal feeling, as a seismologist dealing with natural earthquakes, is one of envy of workers in subjects where it is possible to design experiments at all, and especially where they can be designed in such a way *that the normal equations for the various parameters to be estimated will contain no cross terms*. Nevertheless, I think that some points concerning experimental design are in danger of being overlooked even in the fortunate subjects where design is possible. In particular, the word "randomness" appears to be used in two senses, which are not equivalent. To take an illustration from seismic survey, the experiment would consist in the discharge of an explosion at a known point and the recording of the times of transmission of elastic waves through the ground to a set of recorders at known distances. The choice of the distances is part of the design of the experiment, the problem is to find a pair of parameters a and b as accurately as possible from observations of the times t_r at distances x_r. This will be done by the method of least squares, the equations of condition being

$$a + bx_r = t_r.$$

Now the question of design will be, how should the arbitrary distances x_r be chosen? In practice the intervals are usually made as nearly equal as possible, so that there is a high degree of system in the design. But an advocate of extreme randomness in design would apparently assign all the distances by Tippett's numbers or some analogous device. A possible consequence of this would be that they would all be nearly equal, and it would be impossible to determine either a or b without external evidence concerning the value of the other. Or they might all be concentrated near two values, then a and b could be determined, but there would be no means of testing whether the time of transmission is adequately represented by a linear formula—there are cases in practice where a square or a cube term needs to be included, though the linear form is usually adequate. Thus randomness in design may in some circumstances lead to the loss of information that would be provided by a systematic design.

On the other hand it is claimed that a systematic design may lead to an underestimate of the uncertainty, and it is here that the ambiguity of the use of "randomness" enters. The validity of the method of least squares rests on the hypothesis of randomness in the sense that parameters a and b exist such that, given their values and the standard error, the probabilities of the residuals at all the distances used are independent. This may be true without the distances themselves being assigned at random. Indeed, if the ground is not level, the variations of height will affect the times of transmission for some waves; random-

ness of design will then permit several observations to be near the same distance, and the probabilities of the residuals, given the distances, will not be independent. If in an actual case the variation of level was harmonic, and the recorders were at uniform intervals so as always to be on the crests, there would be a systematic error in the least squares solution; but this can in this problem be calculated and allowed for. In practice, however, variations of level are usually irregular, and independence of the errors is best achieved by using uniform intervals so as to minimize the increase of uncertainty due to possible correlations between neighbouring distances. This does not exclude the possibility that a closer approximation might be attempted by including a term proportional to the height in the law assumed; but that would still not eliminate possible effects of local variation in the velocities immediately below the recording instruments, and the systematic design would still approximate to the condition of independence of the errors better than the random one would.

My excuse for mentioning this problem in a mainly biological journal is that it is one that I am fairly familiar with, and that it seems to illustrate two points that have not been adequately analysed in the biological literature. The first is the absolute necessity of distinguishing between the probabilities of the same proposition assessed on different data. Before designing the above experiment we may have only the vague knowledge of a and b suggested by similar experiments elsewhere. If we are to make a systematic design we do know something about what the relations between the intervals will be; if we are to make a random one we do not. But in either case, as soon as the experiment is designed, the distances are definitely known in both cases, and are not random any longer. The knowledge of the possible types of design that might have been adopted is no longer relevant in either case, since we know the particular design that *has* been adopted; it is now part of the data of the problem. The estimates will in either case be made using the actual distances as data, not some aggregate of the possible distances that might have occurred with that method of design. The hypothesis of the randomness of the residuals, which is needed for the validity of the method of least squares, has nothing to do, intrinsically, with the intended randomness of the original design; in this case the latter randomness would in practice often vitiate the validity of the former, while in any case increasing the uncertainty of the estimates.

The second point is that any method of estimation presupposes a law of error or chance containing adjustable parameters, such that if these were known the probabilities of various experimental results are assigned—in other words, that the likelihood, given these parameters, has a definite value. Now I maintain that this is always what logicians call a *considered* proposition, not an *asserted* one. There is no proof in any actual case that no parameters other than those so far thought of will ever be required; any theory of significance tests contemplates the possibility that others may be needed. But if a new parameter, whose value

is not yet known, may be relevant, then the law of error, given those already considered, is not unique and the likelihood has not a definite value. In the seismic survey, we do not assert that the relation between distance and time of travel is linear; we merely consider it and use it until we find some definite evidence against it in a particular case. But we need the information that will find out if it is wrong; and this must be done by making *sure* that the presence of higher powers can be tested. This is done by the systematic design. The random design would leave it to chance, and abandon the attempt to test the law if it should give an arrangement with all the distances concentrated near two values. It would be pleasant to be able to say that we already know what parameters can be relevant; but we have to face the fact that we do not. What we can do is to provide a procedure for testing them as they arise; but this necessarily implies that we must proceed from the hypothesis involving fewer adjustable parameters to the one involving more. The hypothesis that a limited number of parameters is adequate must be *considered* in any case, but that is not the same as saying that no others will ever be needed. Estimates, however, are always subject to the hypothesis that no parameters other than those explicitly considered are relevant. There appears to be no escape from this dilemma; we can only admit it, state the hypothesis explicitly, and say that the results are the best that we can give in the actual state of our knowledge at the time. Confidence in it will depend mainly on the failure to find evidence for other parameters that might conceivably have been relevant to the observations already available.

Let us now consider how this may apply to problems of sampling a population and the design of agricultural experiments. In the former case the population consists of several different types, of numbers $\nu_1, \nu_2, \ldots, \nu_p$, total ν. The numbers of these types in our sample are $n_1, n_2, \ldots, n_p$, total n. It is not in our power to change the ν_r, and practical considerations may fix n. But $n_1, n_2, \ldots$ are at our disposal. The question is, how should they be chosen? Neyman (1934) has shown that if the standard errors of the respective types are $\sigma_1, \sigma_2, \ldots$, and the means found for the types are x_r, then provided that the ν_r are known the most accurate estimate of the mean for the entire population would be got by taking the n_r in proportion to $\nu_r\sigma_r$, and the estimate as $\Sigma\nu_r x_r/\nu$. If all the σ_r are equal, the n_r should be taken in proportion to the ν_r, and the best estimate of the mean for the population will be the mean over the sample, $\Sigma n_r x_r/n$, irrespective of the differences between different types. The standard error also is independent of these differences. It reduces simply to a matter of estimating the mean for each type from a sample of that type, and its uncertainty is determined only by the variation within types; and the mean for the whole population is the mean of the type means, weighted in proportion to the ν_r, which are in proportion to the n_r. Now this is a highly systematic method of sampling. The random method would be to choose n individuals from the entire population by selecting from a directory by some set of "random numbers". The numbers of the types in the sample

would then depart from proportionality, on account of the usual sampling errors, and if the mean of the sample is taken as the estimate of the mean of the population, it will have an additional error on account of the contributions from the products of the type differences and the sampling errors. It is true that the standard error of the result so obtained is correctly estimated from the mean square variation within the sample as a whole, but it will be larger than that of the mean of a proportional sample. If instead of taking the mean of the sample as our estimate of the mean of the population, thus weighting the type means according to n_r, we weight them according to ν_r, we again recover an estimate that is independent of the variation of the type means and therefore has a higher accuracy, approaching that for the proportional sample.

To apply this method it is necessary that the ν_r should be known. If they are unknown the only way of estimating them would be by taking a random sample from the whole population, and their most probable values would be in the ratios of the n_r. Then the mean of the sample would be the best estimate of the population mean, and its standard error would be correctly found from the variation in the whole sample. The point is that information about the ν_r is relevant to the estimate of the population mean and its standard error, given the sample; and this will be our actual position, since our problem is in practice to make inferences about the population given the sample. If we do not know the population numbers apart from what the sample can tell us about them, there is no more to be said; but if we do know them and omit to use them we are discarding valuable information and introducing avoidable error into the estimate of the mean. The conclusions drawn from a given sample are not the same if the ν_r are also part of the data as they are if the ν_r are unknown.

On the other hand it is claimed that the method of random sampling is unbiased. Now the word "bias" also seems to be capable of several interpretations. If the standard deviations within types are different, but their numbers are known, it has been seen that the best estimate will be obtained by deliberately introducing a departure from proportionality into the sampling and allowing for it afterwards by weighting according to ν_r. It might be said, not unfairly, that any procedure that gives an estimate different from the best one using the whole of the available data is biased. It is true that random sampling at the outset is designed to give every possible sample of given size an equal chance of being the actual sample; but when the sample has actually been taken, that is the sample and the intentions of the designer are no longer relevant. If we know that it is not in fact proportional we are entitled to allow for the departure and shall gain accuracy by doing so. The fact that the use of the sample mean as the estimate might give an error with either sign in a future experiment does not imply that in a particular case it is equally likely to have given one with either sign. It is sometimes said that by taking a random sample we avoid dangers that might arise through a systematic sample not being in proportion. Thus Yule & Kendall

(1937, p. 349) mention the problem of sampling in a planned town, where, if we choose houses at equal intervals equal to the length of the blocks, we should get a sample containing either no corner houses or nothing else. But here the difficulty can be avoided either by choosing an interval that is prime to the number of houses in the blocks, or by deliberately taking samples of corner and other houses separately. It is true that such a sample does not give all combinations of houses an equal chance of being chosen, but that appears to be a merit. It excludes the possibility that might still arise in random sampling, that we should miss all the corner houses, and it does not necessarily invalidate the hypothesis of the randomness of the errors. In this case it is not necessary that the houses already chosen should be irrelevant to the identity of the next; it is sufficient that they should be irrelevant to the amount of the departure of the next from the mean of its type, which is a much less stringent condition, and the considerations mentioned in relation to the problem of seismic survey will apply. My point is that if we omit to make use of relevant information that is in our possession, we are making avoidable errors, and in this problem the possible difference between corner and other houses is a known piece of information, and the best procedure is to design the work so as to determine it as accurately as possible and not leave it to chance whether it can be determined at all. It is possible in any case that the hypothesis of independence of the errors may be wrong; no absolute provision against that can be made, and it may lead to unexpected discrepancies. But that is dealt with by recognizing that the hypothesis is a considered proposition, and that the argument is inductive and not deductive; it is not dealt with by including an estimable error that has nothing to do with it.

I have been puzzled by the recent criticisms of Fisher by "Student" (1937). Reading "Student's" paper and Fisher's *Design of Experiments* I find myself in almost complete agreement with both; and I should therefore have expected them to agree with each other. On the matter of the half-drill strip I have no opinion whatever, and leave that to specialists. But it seems to me that "Student" is wrong in regarding Fisher as an advocate of extreme randomness, and possibly Fisher has not sufficiently emphasized the amount of system in his methods. Consider an arrangement intended to compare five varieties or treatments in three 5×5 squares. A die-hard randomist might proceed by numbering the plots in order, 1 to 75, and throw a die till numbers other than sixes have occurred 75 times. If the nth throw, neglecting sixes, gives a five, he will assign the nth plot to the fifth variety, and so on. He might carry out an analogous procedure with some series of random numbers. The result might be that some variety might not appear at all in one square; there would usually be many cases where the same variety occurred two or three times in the same row or column. Different varieties would not occur equally often in the design, and the comparisons between them would have different weights. It is excessively improbable that the result would be three Latin squares. When Fisher makes each variety occur

equally often he introduces one systematic restriction, when he applies this to each square separately he introduces another, when he makes it occur just once in every row and every column he introduces others. The complete random procedure has never, I think, been adopted in such a case. It may be too absurd to have even been considered; but statement of it shows how far Fisher has gone in the direction advocated by "Student". The latter does in fact point out that the Latin square is both random and balanced, but the two properties refer to different features of the design. The systematic features enable the maximum accuracy to be obtained for each variety, consistently with uniformity. But I think that some further attention should be given to the function of the randomizing of rows and columns that is carried out after the main features of the Latin square design are fixed. There would be no point in this if the probabilities of the plot errors, given the varietal, row and column values, were all independent. The hypothesis required by the method of least squares would be satisfied, and the analysis of variance is equivalent to the method of least squares applied to a system where the normal equations contain only diagonal terms. On the other hand a systematic ground effect that upsets the independence of the errors would require the explicit introduction of a new parameter to express it; this would not necessarily enter orthogonally to the others, and the analysis of variance analysis would have to be replaced by a detailed solution by least squares. Now the absence of such terms cannot be guaranteed. To illustrate how they can enter, let us suppose axes of position taken parallel to the sides, with the centre of the square as origin. Then if (x, y) are the coordinates of the centre of a plot, suppose first that the fertility may be expressed in the form

$$F = a_0 + a_1 x + b_1 y + a_2 x^2 + b_2 y^2 + \dots + b_4 y^4.$$

The elimination of rows and columns, in a 5×5 square, allows for and eliminates all powers of x and y separately up to x^4 and y^4; evidently any set of row and column totals could be fitted by choosing the nine coefficients suitably. But neither the design of the square nor the analysis of variance technique would deal with a term like xy, which might be present. Whatever its coefficient may be, it will contribute nothing to the row and column totals; but Σxy for all plots of a particular variety will not in general vanish—indeed it is possible for xy to have the same sign for four of them, vanishing for the other. Thus the presence of such a term will make a contribution to the estimates of the varietal differences unless provision is made against it. Now this is not merely a theoretical danger. If terms in x^2 and y^2 are relevant it will be only for one particular pair of directions of the sides that the coefficient of xy in the fertility will vanish, and it must be presumed that if there is reason to attend to x^2 and y^2 we should also attend to xy. The most accurate way of doing so would be to introduce it explicitly into the equations of condition and estimate it by the method of least squares. In estimating the uncertainty allowance would have to be made for the fact that an additional parameter has been estimated for every square. But since it will not

enter orthogonally to the varietal differences the labour of calculation would be greatly increased. Again, if x^3 and y^3 are relevant, there is reason to expect x^2y and xy^2 to be, and so on. The analysis would become prohibitively difficult.

Now this is where the use of randomization comes in. In a single 5×5 Latin square it is impossible for Σxy to vanish for every variety. (I do not know whether it would be possible for larger squares, but even if it is, higher product terms would require attention.) But it is possible to arrange the design so that, without evaluating the contributions from xy and analogous terms explicitly, they will contribute independently to the totals. This is done by repeating the square, using separate acts of randomization for each square. It is important to notice that the object would not be achieved if the first square was simply copied. If this was done, the values of Σxy for each treatment in one square, and for one treatment in each of the others, would determine the values for all the other treatments in the other two squares, and the hypothesis of independence would be wrong. The separate randomizations ensure that, with three replications, the contributions of the xy and similar terms to the varietal totals are each the sum of three independent components. Omission to evaluate them explicitly may sacrifice information that could possibly be recovered, and thereby lead to loss of accuracy in estimating the main effects; but so far as that is a criticism it does not apply here to the randomization but to the analysis of variance technique, which will be valid in these circumstances only if some artificial device is introduced to convert the systematic disturbance into one that can legitimately be treated as random. The real justification of using this technique rather than a complete least squares solution, taking into account xy and possibly higher product terms, is practical convenience. The justification of the separate randomization in relation to it is that it prevents differences from being interpreted as due to varietal or treatment differences when they are really due to the accumulation of neglected ground effects. To advocate a least squares solution for the neglected terms may well be analogous to saying that an investigator should carry out all computations to seven figures when he knows quite well that the second will be uncertain in any case.

It seems to me that Fisher sums up the situation very well in his advice to balance or eliminate the larger systematic effects as accurately as possible and randomize the rest. It provides against two dangers: first, that the main object of the work may be lost by an imperfect estimate of the larger effects, especially through great departures of the normal equations from orthogonality; and secondly, that a neglected term, small in any one observation, may mount up when many are combined. But what is worth balancing, what is worth randomizing, at the cost of a small increase of uncertainty, but not worth definitely eliminating, and what is random anyhow, must be a matter of the particular problem. The only rule is to attend to actual conditions and the types of systematic variation likely to arise in them. I should see no point, for instance, in an

elaborate sampling of people by a fixed list of random numbers when a list in alphabetical order is available; if uniform intervals in such a list do not give a random sample in the sense of the independence of the errors I do not know what will. The only systematic variation that could bias the mean would be a periodicity in terms of the position in the table, the wave-length being an exact sub-multiple of the interval chosen; and this seems too remote a possibility to need much consideration. But a similar difficulty might arise in relation to one experimental layout mentioned by "Student", namely

$$A\ B\ B\ A\ A\ B\ B\ A.$$

Here, as "Student" indeed remarks, the estimated difference between A and B would be biased if there was a harmonic variation of fertility whose period is an odd multiple of the width of a quartet. He dismisses this as a not particularly likely occurrence. But land is often ridged in just this way, and an unwary designer might easily lay out his experiment in such a way that the difference sought could not be separated from the periodic ground effect. In such land I should say that the harmonic effect is likely to be the dominant ground effect and that the best treatment would be to introduce it explicitly into the equations of condition, and lay out the experiment so that the possible phases are uniformly distributed both for A and B.

In the 5×5 Latin square, there is no *logical* reason why we should stop before an expression of possible ground effects by an expression of twenty-five terms with adjustable coefficients; then there would be no means left for separating treatment effects from ground ones. Equally there seems to be none for not treating the whole variation in terms of block mean, treatments, and random error. The place to draw the line between these two extremes cannot be decided by pure theory; it is indicated by previous experience, which has indicated that some types of ground effect are habitually significant, others occasionally significant, and others legitimately treated as random. What Fisher's randomization does is really to combine members of the second class in such a way that they may be included in the third. But extreme randomization would also try to randomize the first class, and the disadvantages of this seem far to outweigh any possible advantage.

However, I must now return to a subject where the normal equations are never orthogonal; where the normal law of error never holds; and where it is impossible to separate the ground effects in different regions from possible systematic differences between the observers inhabiting those regions.

REFERENCES

Neyman, J. (1934). *J.R. Statist. Soc.* **97**, 577–80.

"Student" (1937). *Biometrika*, **29**, 363–79.

Yule, G. U. & Kendall, M. G. (1937). *Introduction to the Theory of Statistics.* London: Griffin.

A NOTE ON NORMAL CORRELATION*

By E. J. G. PITMAN
University of Tasmania

In a recent paper Finney (1938) discussed the distribution of the ratio of estimates of the two variances in a sample from a normal bi-variate population, and showed how a test for significance could be applied when the population correlation coefficient is known. He also showed how the test may be adapted when only a sample estimate of this correlation is available, but this adaptation is not completely satisfactory. This note shows how, by using a different distribution, an exact test for significance can be obtained, and fiducial limits for the ratio of the population variances determined, when the population correlation coefficient is unknown.

Suppose that x, y are normally correlated variables with probability function

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-a)^2}{\sigma_1^2}-2\rho\frac{(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\},$$

and that $$x_s, y_s \qquad (s = 1, 2, \ldots, n)$$

are n pairs of observed values of x, y. Write

$$\bar{x} = \Sigma x_s/n, \qquad \bar{y} = \Sigma y_s/n,$$

$$S_1 = \Sigma(x_s-\bar{x})^2, \qquad S_2 = \Sigma(y_s-\bar{y})^2,$$

$$r = \frac{\Sigma(x_s-\bar{x})(y_s-\bar{y})}{\sqrt{\{\Sigma(x_s-\bar{x})^2 \cdot \Sigma(y_s-\bar{y})^2\}}}.$$

Since $$-\frac{1}{2(1-\rho^2)}\left[\frac{(x-a)^2}{\sigma_1^2}-2\rho\frac{(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]$$

$$= -\frac{1}{2}\left[\frac{1}{2(1+\rho)}\left\{\frac{x-a}{\sigma_1}+\frac{y-b}{\sigma_2}\right\}^2+\frac{1}{2(1-\rho)}\left\{\frac{x-a}{\sigma_1}-\frac{y-b}{\sigma_2}\right\}^2\right],$$

it is evident that $$u = \frac{x}{\sigma_1}+\frac{y}{\sigma_2}, \quad v = \frac{x}{\sigma_1}-\frac{y}{\sigma_2}$$

are *independent* normal variables with variances $2(1+\rho)$, $2(1-\rho)$ respectively.

Thus, if $$u_s = \frac{x_s}{\sigma_1}+\frac{y_s}{\sigma_2}, \quad v_s = \frac{x_s}{\sigma_1}-\frac{y_s}{\sigma_2}, \qquad (s = 1, 2, \ldots, n),$$

the n pairs of numbers u_s, v_s constitute a sample from a normal bi-variate

* [This paper and that which follows by W. A. Morgan were received for publication at about the same date. The subject of both was suggested by D. J. Finney's paper, but the two treatments and applications of the improved solution of his problem, when ρ is unknown, follow quite different lines. Ed.]

population with zero correlation. Hence the distribution of their correlation coefficient,

$$R = \frac{\Sigma(u_s - \bar{u})(v_s - \bar{v})}{\sqrt{\{\Sigma(u_s - \bar{u})^2 \,.\, \Sigma(v_s - \bar{v})^2\}}},$$

is known.

$$R = \frac{\Sigma\left(\frac{x_s - \bar{x}}{\sigma_1} + \frac{y_s - \bar{y}}{\sigma_2}\right)\left(\frac{x_s - \bar{x}}{\sigma_1} - \frac{y_s - \bar{y}}{\sigma_2}\right)}{\sqrt{\left\{\Sigma\left(\frac{x_s - \bar{x}}{\sigma_1} + \frac{y_s - \bar{y}}{\sigma_2}\right)^2 . \Sigma\left(\frac{x_s - \bar{x}}{\sigma_1} - \frac{y_s - \bar{y}}{\sigma_2}\right)^2\right\}}}$$

$$= \frac{\frac{S_1}{\sigma_1^2} - \frac{S_2}{\sigma_2^2}}{\sqrt{\left\{\left(\frac{S_1}{\sigma_1^2} + \frac{S_2}{\sigma_2^2} + \frac{2r\sqrt{(S_1 S_2)}}{\sigma_1 \sigma_2}\right)\left(\frac{S_1}{\sigma_1^2} + \frac{S_2}{\sigma_2^2} - \frac{2r\sqrt{(S_1 S_2)}}{\sigma_1 \sigma_2}\right)\right\}}}$$

$$= \frac{\frac{S_1}{\sigma_1^2} - \frac{S_2}{\sigma_2^2}}{\sqrt{\left\{\left(\frac{S_1}{\sigma_1^2} + \frac{S_2}{\sigma_2^2}\right)^2 - 4r^2 \frac{S_1 S_2}{\sigma_1^2 \sigma_2^2}\right\}}}.$$

Putting $\qquad \omega = \sigma_1^2/\sigma_2^2, \quad w = S_1/S_2,$

we have $\qquad R = \dfrac{w - \omega}{\sqrt{\{(w + \omega)^2 - 4r^2 w\omega\}}}.$

The values of w and r are given by the sample, hence, from the known distribution of R, fiducial limits for ω can be determined. If we merely wish to test whether the values of S_1 and S_2 are significantly different, i.e. if we wish to test the null hypothesis $\omega = 1$, we simply insert this value of ω and test for significance the corresponding value of the correlation coefficient R. In determining fiducial limits for ω the arithmetic is a little easier if we make use of "Student's" z-distribution, or Fisher's t-distribution. We know that R^2 has a $B(\frac{1}{2}, \frac{1}{2}n - 1)$ distribution, and that

$$\frac{R}{\sqrt{(1 - R^2)}} = \frac{w - \omega}{\sqrt{\{4(1 - r^2)\, w\omega\}}}$$

is distributed like "Student's" z for a sample of $n - 1$, while

$$t = \frac{R\sqrt{(n-2)}}{\sqrt{(1 - R^2)}} = \frac{(w - \omega)\sqrt{(n-2)}}{\sqrt{\{4(1 - r^2)\, w\omega\}}}$$

is distributed like Fisher's t with $n - 2$ degrees of freedom.

If α is any given number less than 1, we can determine k such that

$$P\{|t| \leqq k\} = \alpha.$$

The inequality $|t| \leqq k$ is equivalent to

$$(w - \omega)^2 \leqq \frac{4(1 - r^2)\, k^2 w\omega}{n - 2},$$

i.e. $\qquad \omega^2 - 2Kw\omega + w^2 \leqq 0,$

where $$K = 1 + \frac{2(1-r^2)k^2}{n-2};$$

hence $$P\{w(K - \sqrt{(K^2-1)}) \leqq \omega \leqq w(K + \sqrt{(K^2-1)})\} = \alpha.$$

In the example given by Finney (1938, p. 192),

$$n = 173, \quad w = \frac{S_1}{S_2} = \frac{(5{\cdot}209)^2}{(4{\cdot}766)^2}, \quad r = 0{\cdot}878.$$

Substituting these values in the expression for t, and putting $\omega = 1$, we obtain $t = 2{\cdot}902$. As n is large the distribution of t is approximately normal with zero mean and unit standard deviation, and therefore the value of t is significant at the 1% level. It may be noted that in this particular case the exact test gives less significance than Finney's.

For comparison with the case of uncorrelated normal variables we may note that if $\rho = 0$,

$$\frac{w}{w+\omega}$$

has a $B\{\frac{1}{2}(n-1), \frac{1}{2}(n-1)\}$ distribution, and therefore

$$\frac{(w-\omega)^2}{(w+\omega)^2}$$

has a $B\{\frac{1}{2}, \frac{1}{2}(n-1)\}$ distribution. As shown above, for any value of ρ,

$$R^2 = \frac{(w-\omega)^2}{(w+\omega)^2 - 4r^2 w\omega}$$

has a $B\{\frac{1}{2}, \frac{1}{2}(n-2)\}$ distribution.

The same method gives very easily* the distribution required for the estimation of ρ when the standard deviations of x and y are the same, $\sigma_1 = \sigma_2 = \sigma$.

$$u = x + y, \quad v = x - y$$

are then independent normal variables with variances $2\sigma^2(1+\rho)$, $2\sigma^2(1-\rho)$;

hence $$\frac{\Sigma(u_s - \bar{u})^2}{2\sigma^2(1+\rho)}, \quad \frac{\Sigma(v_s - \bar{v})^2}{2\sigma^2(1-\rho)}$$

are independent chance variables each distributed like χ^2 with $n-1$ degrees of freedom, and so the distribution of their quotient

$$w = \frac{1-\rho}{1+\rho} \cdot \frac{\Sigma(u_s - \bar{u})^2}{\Sigma(v_s - \bar{v})^2}$$

is known; $w/(1+w)$ has a $B\{\frac{1}{2}(n-1), \frac{1}{2}(n-1)\}$ distribution. The expression for w is

$$w = \frac{1-\rho}{1+\rho} \cdot \frac{1+r}{1-r},$$

where $$r = \frac{2\Sigma(x_s - \bar{x})(y_s - \bar{y})}{\Sigma(x_s - \bar{x})^2 + \Sigma(y_s - \bar{y})^2}.$$

* Cf. De Lury (1938), where the same problem is dealt with by a different method.

If x and y have the same mean value a as well as the same standard deviation σ, the mean value of r is 0, and therefore

$$\frac{\Sigma v_s^2}{2\sigma^2(1-\rho)}$$

is distributed like χ^2 with n degrees of freedom. Thus the distribution of the quotient,

$$w = \frac{1-\rho}{1+\rho} \cdot \frac{\Sigma(u_s - u)^2}{\Sigma v_s^2},$$

is known; $w/(1+w)$ has a $B\{\tfrac{1}{2}(n-1), \tfrac{1}{2}n\}$ distribution. The expression for w is

$$w = \frac{1-\rho}{1+\rho} \cdot \frac{1+r}{1-r},$$

where

$$r = \frac{2\Sigma(x_s - \xi)(y_s - \xi)}{\Sigma(x_s - \xi)^2 + \Sigma(y_s - \xi)^2},$$

$$\xi = \Sigma(x_s + y_s)/2n.$$

The procedure in this case corresponds exactly to the treatment of fraternal correlation by the analysis of variance.

REFERENCES

FINNEY, D. J. (1938). *Biometrika*, **30**, 190–2.

DE LURY, D. B. (1938). *Ann. Math. Statist.* **9**, 2, 149–51.

A TEST FOR THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE TWO VARIANCES IN A SAMPLE FROM A NORMAL BIVARIATE POPULATION

By W. A. MORGAN

Department of Statistics, University College, London

1. Derivation of likelihood ratio test

In a paper published in a recent issue of this Journal D. J. Finney (1938) considered the following questions. A sample of n pairs of variables (x, y) has been drawn from the bivariate normal distribution whose probability law is

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho_{12}^2)}}\exp\left[-\frac{1}{2(1-\rho_{12}^2)}\left\{\left(\frac{x-\xi_1}{\sigma_1}\right)^2 - 2\rho_{12}\frac{(x-\xi_1)(y-\xi_2)}{\sigma_1\sigma_2} + \left(\frac{y-\xi_2}{\sigma_2}\right)^2\right\}\right], \quad \text{......(1)}$$

then

(i) What is the probability law of the ratio, ω, where

$$\omega = s_1'/s_2' \quad \text{......(2)}$$

and

$$(s_1')^2 = \sum_i (x_i - \bar{x})^2/(n-1), \quad (s_2')^2 = \sum_i (y_i - \bar{y})^2/(n-1)\,? \quad \text{......(3)}$$

(ii) Could this ratio be used as a criterion to test the hypothesis that $\sigma_1 = \sigma_2$?

Using a more direct method he was first able to confirm a previous result of Bose (1935), giving the probability distribution of ω in the case where $\sigma_1 = \sigma_2$, and hence to show that the chance that ω exceeds a given value, say Ω, could be obtained from the *Tables of the Incomplete Beta Function* (1934), using the relation

$$P\{\omega > \Omega\} = I_x\left(\frac{n-1}{2}, \frac{n-1}{2}\right), \quad \text{......(4)}$$

where

$$x = \frac{1}{2}\left(1 - \frac{\Omega - \Omega^{-1}}{\sqrt{\{(\Omega + \Omega^{-1})^2 - 4\rho_{12}^2\}}}\right). \quad \text{......(5)}$$

Since the probability expression (4) is dependent on the population correlation ρ_{12}, which will be in general unknown, Finney pointed out that the ratio ω was not altogether a satisfactory criterion to use in testing the hypothesis $\sigma_1 = \sigma_2$, but he put forward a possible method of getting round this difficulty.

The question arises as to whether there is not some other more suitable criterion for testing whether the variances are equal, whose sampling distribution will be independent of ρ_{12} and of any other parameters whose values are not specified by the hypothesis itself. The likelihood ratio method of approach

of Neyman and Pearson may be followed: this method has proved of service in a number of instances where the appropriate criterion was not immediately obvious. Summarized briefly, it involves the following steps.

(*a*) A specification of the set of admissible hypotheses. In the present case these will be defined by the joint probability law of the n pairs of observations,

$$p(x_1, \ldots, x_n; y_1, \ldots, y_n \mid \xi_1, \xi_2, \sigma_1, \sigma_2, \rho_{12})$$
$$= \{2\pi\sigma_1\sigma_2\sqrt{(1-\rho_{12}^2)}\}^{-n} \exp\left[-\frac{n}{2(1-\rho_{12}^2)}\left\{\left(\frac{\bar{x}-\xi_1}{\sigma_1}\right)^2 - 2\rho_{12}\frac{(\bar{x}-\xi_1)(\bar{y}-\xi_2)}{\sigma_1\sigma_2}\right.\right.$$
$$\left.\left. + \left(\frac{\bar{y}-\xi_2}{\sigma_2}\right)^2 + \frac{s_1^2}{\sigma_1^2} - \frac{2\rho_{12}r_{12}s_1s_2}{\sigma_1\sigma_2} + \frac{s_2^2}{\sigma_2^2}\right\}\right], \qquad \ldots (6)$$

where $\qquad -\infty \leq \xi_1, \xi_2 \leq \infty, \quad 0 \leq \sigma_1, \sigma_2 \leq \infty, \quad -1 \leq \rho_{12} \leq 1.$

In this expression $\bar{x}$ and $\bar{y}$ are the sample means, r_{12} the sample correlation coefficient, and

$$s_1^2 = \sum_i (x_i - \bar{x})^2/n, \quad s_2^2 = \sum_i (y_i - \bar{y})^2/n. \qquad (7)$$

(*b*) A determination of those values of the five unknown parameters, as functions of the observations, which jointly maximize the expression (6). The solution is known to be obtained when

$$\xi_1 = \bar{x}, \quad \xi_2 = \bar{y}, \quad \sigma_1 = s_1, \quad \sigma_2 = s_2, \quad \rho_{12} = r_{12}. \qquad (8)$$

The maximum value of (6) is thus

$$p_1(\max) = \{2e\pi s_1 s_2\sqrt{(1-r_{12}^2)}\}^{-n}. \qquad (9)$$

(*c*) A specification of the hypothesis tested. This hypothesis assumes that the probability law is of the form

$$p(x_1, \ldots, x_n; y_1, \ldots, y_n \mid \xi_1, \xi_2, \sigma, \rho_{12}),$$

where the function is obtained by putting $\sigma_1 = \sigma_2 = \sigma$ in (6).

(*d*) A determination of the values of the four unknown parameters ξ_1, ξ_2, σ, and ρ_{12} which maximize this expression. These values may be shown to be

$$\xi_1 = \bar{x}, \quad \xi_2 = \bar{y}, \quad \sigma = \sqrt{\{\tfrac{1}{2}(s_1^2 + s_2^2)\}}, \quad \rho_{12} = 2r_{12}s_1s_2/(s_1^2 + s_2^2). \quad \ldots\ldots (10)$$

The maximum value of the probability function defined in (*c*) then becomes

$$p_2(\max) = \{e\pi\sqrt{[(s_1^2+s_2^2)^2 - 4r_{12}^2 s_1^2 s_2^2]}\}^{-n}. \qquad \ldots\ldots (11)$$

(*e*) The likelihood ratio criterion is then

$$\lambda = \frac{p_2(\max)}{p_1(\max)} = \left\{\frac{4s_1^2s_2^2(1-r_{12}^2)}{(s_1^2+s_2^2)^2 - 4r_{12}^2s_1^2s_2^2}\right\}^{\frac{1}{2}n} \qquad \ldots\ldots (12a)$$
$$= \left\{1 - \frac{(s_1^2-s_2^2)^2}{(s_1^2+s_2^2)^2 - 4r_{12}^2s_1^2s_2^2}\right\}^{\frac{1}{2}n}. \qquad \ldots\ldots (12b)$$

(*f*) The hypothesis tested becomes less and less likely as λ moves from 1 to 0. To complete the test it is necessary to know the sampling distribution of λ, or of a single valued function of λ, when the hypothesis tested is true.

It will be seen at once that the test differs from Finney's because the criterion,

unlike his ω, is a function of r_{12} as well as s_1^2 and s_2^2. The meaning of the criterion which has been picked out by the λ-method becomes clear if we make the following transformation of the original variables:

Write $$x = X + Y, \qquad y = X - Y, \qquad \text{......(13)}$$

so that $$X = \tfrac{1}{2}(x+y), \quad Y = \tfrac{1}{2}(x-y). \qquad \text{......(14)}$$

Then the population variances of x and y may be expressed as functions of the variances and correlation for X and Y, as follows:

$$\left.\begin{aligned} \sigma_1^2 &= \sigma_X^2 + \sigma_Y^2 + 2\rho_{XY}\sigma_X\sigma_Y, \\ \sigma_2^2 &= \sigma_X^2 + \sigma_Y^2 - 2\rho_{XY}\sigma_X\sigma_Y. \end{aligned}\right\} \qquad \text{......(15)}$$

The necessary and sufficient condition that the hypothesis tested is true, or that $\sigma_1 = \sigma_2$, is that
$$\rho_{XY} = 0. \qquad \text{......(16)}$$

Since X and Y are normally correlated variables, the appropriate criterion to test the hypothesis, $\rho_{XY} = 0$, is the sample correlation coefficient between the transformed variables, i.e. r_{XY}. If the hypothesis is true, this coefficient has the well-known probability law

$$p(r_{XY} \mid \rho_{XY} = 0) = \text{constant} \times (1 - r_{XY}^2)^{\frac{1}{2}(n-4)}. \qquad \text{......(17)}$$

Making use of the transformations (14), it is found that

$$r_{XY} = \frac{s_1^2 - s_2^2}{\{(s_1^2 + s_2^2)^2 - 4r_{12}^2 s_1^2 s_2^2\}^{\frac{1}{2}}}. \qquad \text{......(18)}$$

Hence the likelihood criterion of (12b) is seen to be

$$\lambda = \{1 - r_{XY}^2\}^{\frac{1}{2}n}, \qquad \text{......(19)}$$

and as the hypothesis tested becomes less and less likely, $\lambda \to 0$ or $r_{XY}^2 \to 1$. The test may therefore be carried out by (a) referring the r_{XY} of (18) to the probability distribution (17), or (b) alternatively referring

$$t = \frac{r_{XY}\sqrt{(n-2)}}{\sqrt{(1 - r_{XY}^2)}} \qquad \text{......(20)}$$

to "Student's" distribution with degrees of freedom $f = n - 2$, and (c) rejecting the hypothesis when $|r_{XY}|$ or $|t|$ fall beyond the desired probability level. The test, it will be seen, is independent of the unknown correlation ρ_{12} between x and y.

2. THE POWER OF THE TEST

While the probability distribution of r_{XY} is independent of ρ_{12} if the hypothesis ($\sigma_1 = \sigma_2$) is true, the chance that in using the rejection rule of the test we shall detect real differences between σ_1 and σ_2 will depend on the value of ρ_{12}. Suppose that we fix a probability level $r(\alpha)$ such that

$$\alpha = 2\int_{r(\alpha)}^{1} c(1 - r^2)^{\frac{1}{2}(n-4)}\,dr,* \qquad \text{......(21)}$$

where α, for example, equals 0·05.

* c is the constant of the probability law (17).

Then the chance of rejecting the hypothesis that

$$\gamma = \sigma_1/\sigma_2 = 1, \tag{22}$$

when $\gamma = \gamma_1 \neq 1$, is given by the expression

$$P\{r^2_{XY} > r^2(\alpha) \mid \gamma = \gamma_1\} = P\{r^2 > r^2(\alpha) \mid \rho = \rho_1 \neq 0\}$$

$$= 1 - \int_{-r(\alpha)}^{r(\alpha)} p(r \mid \rho = \rho_1)\, dr. \tag{23}$$

This expression Neyman & Pearson (1936) have termed the power of the test of the hypothesis that $\gamma = \sigma_1/\sigma_2 = 1$ with regard to an alternative hypothesis $\gamma = \gamma_1$. In the expression (23), $p(r \mid \rho = \rho_1)$ denotes the general probability law for r in samples from a bivariate normal population, first obtained by R. A. Fisher (1915). It will be seen that a relation similar to (18) holds between ρ_{XY}, $\gamma = \sigma_1/\sigma_2$ and ρ_{12}, namely

$$\rho_{XY} = \frac{\gamma - \gamma^{-1}}{\{(\gamma - \gamma^{-1})^2 + 4(1 - \rho_{12}^2)\}^{\frac{1}{2}}}. \tag{24}$$

Owing to the symmetry of the distribution of r when $\rho = 0$, the power of the test will be the same for alternatives ρ_{XY} and $-\rho_{XY}$, it will therefore be the same for alternatives γ and γ^{-1}. For example, the test is as likely to reject the hypothesis $(\sigma_1 = \sigma_2)$ when $\sigma_1 = 2\sigma_2$ as when $\sigma_1 = \frac{1}{2}\sigma_2$. This is clearly what we should expect, as σ_1 and σ_2 are in no way differentiated.

In Fig. 1 I have shown the power function of the test, taking $\alpha = 0.10$* and sample size $n = 25$, for alternatives $\gamma > 1$, in the three cases $\rho_{12} = 0$, $\rho_{12} = 0.5$, and $\rho_{12} = 0.8$. It will be noticed that the test is more powerful when ρ_{12} is large. This of course follows from (24), since for a given value of γ, ρ_{XY} will be further from zero the nearer $|\rho_{12}|$ is to unity.

The computations were made with the help of F. N. David's recently published *Tables* (1938). The work was simplified by taking ρ_{XY} at convenient values 0·1, 0·2, ..., 0·9, and finding the corresponding values of γ by means of (24).

The table on p. 18 shows, in the columns headed Test (a), the values of the power function computed in this way for this case of $n = 25$ and also for $n = 12$ and $n = 100$.

3. Comparison with Finney's test in the case where ρ_{12} is known

In the case where ρ_{12} is not known, Finney has suggested that his test criterion ω might be used by making a double appeal to significance levels on the lines proposed in another case by Hirschfeld (1937). It does not, however, appear easy to determine numerically the power of the resulting test. In the case where ρ_{12} is known, it may be shown that the likelihood ratio criterion now becomes

$$\lambda = \left\{1 + \frac{(s_1 - s_2)^2}{2s_1 s_2(1 - r_{12}\rho_{12})}\right\}^{-n} = \left\{1 + \frac{(\omega - 1)^2}{2\omega(1 - r_{12}\rho_{12})}\right\}^{-n}. \qquad \ldots\ldots(25)$$

* This means that the hypothesis is to be rejected when r_{XY} falls beyond the 5% level in either tail of the distribution (17), or for the case $n = 25$, when $|r_{XY}| > 0.3365$.

This is not solely a function of Finney's criterion ω, since it depends again on the sample correlation r_{12}. The form of this expression is interesting, as it shows that if ρ_{12} is known and differs from zero, in view of the correlation which exists between s_1, s_2 and r_{12},* the sample values of all three are relevant in examining the hypothesis that $\sigma_1 = \sigma_2$. I have not succeeded in determining the sampling distribution of the λ of (25), or of any single-valued function of λ.

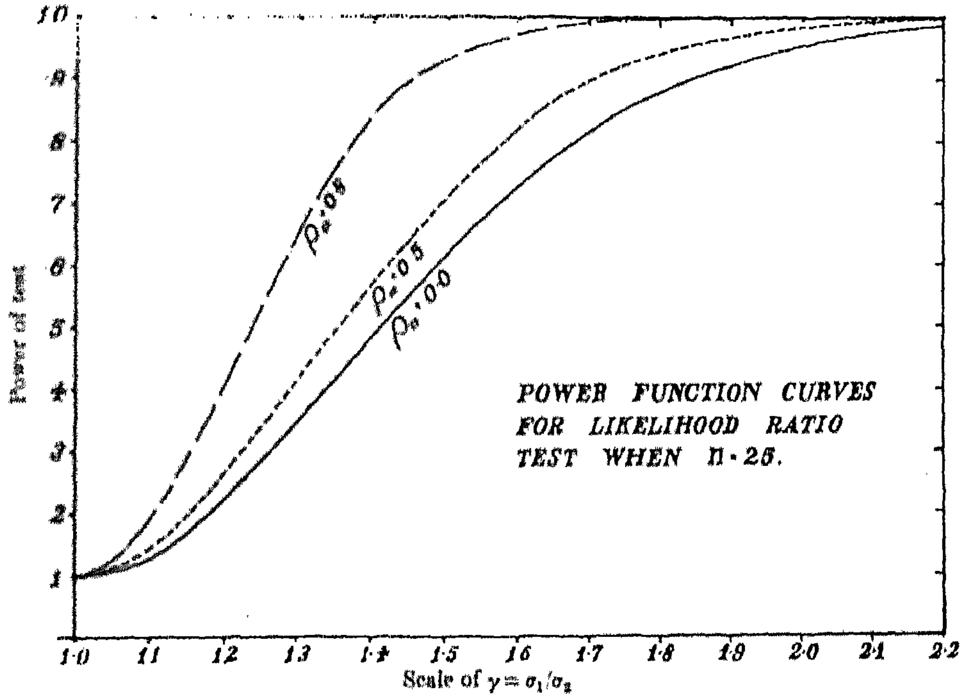


Fig. 1.

In the case of ρ_{12} known, it is, however, possible to compare Finney's test (involving ρ_{12}) with the likelihood test appropriate in the case where ρ_{12} is unknown, but still of course applicable when it is known. The power function for Finney's test may be computed as follows.

When using this test the hypothesis that $\gamma = 1$ is rejected if $\omega > \Omega$ or if $\omega < \Omega^{-1}$, where Ω is a constant chosen by using relations (4) and (5) so that

$$\int_{\Omega}^{\infty} p(\omega \mid \gamma = 1)\, d\omega = \tfrac{1}{2}\alpha. \qquad \ldots\ldots(26)$$

The power of the test with regard to an alternative hypothesis, $\gamma \neq 1$, is therefore given by

$$F(\gamma) = 1 - P\{\Omega^{-1} < \omega < \Omega\} = \int_0^{\Omega^{-1}} p(\omega \mid \gamma)\, d\omega + \int_{\Omega}^{\infty} p(\omega \mid \gamma)\, d\omega, \qquad \ldots\ldots(27)$$

where

$$p(\omega \mid \gamma) = \text{constant} \times (1-\rho_{12}^2)^{\frac{1}{2}(n-1)} \frac{\gamma^2+\omega^2}{\gamma\omega^2} \left\{ \left(\frac{\gamma^2+\omega^2}{\gamma\omega} \right)^2 - 4\rho_{12}^2 \right\}^{-\frac{1}{2}n}. \qquad \ldots\ldots(28)$$

* See K. Pearson (1913).

The above integrals may be transformed to give

$$F(\gamma) = 1 - I_{x_1}\left(\frac{n-1}{2}, \frac{n-1}{2}\right) + I_{x_2}\left(\frac{n-1}{2}, \frac{n-1}{2}\right) \qquad \ldots \quad (29)$$

where $\quad x_1 = \frac{1}{2}[1 + (\gamma\Omega - \gamma^{-1}\Omega^{-1})/\sqrt{\{(\gamma\Omega - \gamma^{-1}\Omega^{-1})^2 + 4(1-\rho_{12}^2)\}}]$

and $\quad x_2 = \frac{1}{2}[1 - (\gamma^{-1}\Omega - \gamma\Omega^{-1})/\sqrt{\{(\gamma^{-1}\Omega - \gamma\Omega^{-1})^2 + 4(1-\rho_{12}^2)\}}] \qquad (30)$

so that values of the power function may be calculated using the *Tables of the*

Comparison of power functions of (a) Likelihood test (based on r_{XY}), (b) Finney's test (based on w), when ρ_{12} is known

Case $\rho_{12}=0{\cdot}0$		$n=12$		$n=25$		$n=100$	
$\gamma=\sigma_1/\sigma_2$	ρ_{XY}	Test (a)	Test (b)	Test (a)	Test (b)	Test (a)	Test (b)
1·00	0	0·100	0·100	0·100	0·100	0·100	0·100
1·11	0·1	0·116	—	0·138	0·138	0·262	0·262
1·22	0·2	0·167	—	0·254	0·254	0·642	0·642
1·36	0·3	0·254	—	0·437	0·438	0·922	0·922
1·53	0·4	0·379	0·382	0·652	0·653	0·979	
2·00	0·6	0·710	0·715	0·953	0·954		
3·00	0·8	0·965	—	1·000	1·000		

Case $\rho_{12}=0{\cdot}5$		$n=12$		$n=25$		$n=100$	
$\gamma=\sigma_1/\sigma_2$	ρ_{XY}	Test (a)	Test (b)	Test (a)	Test (b)	Test (a)	Test (b)
1·00	0	0·100	0·100	0·100	0·100	0·100	0·100
1·09	0·1	0·116	0·116	0·138	0·138	0·262	0·261
1·19	0·2	0·167	0·166	0·254	0·253	0·642	0·643
1·31	0·3	0·254	—	0·437	0·437	0·922	0·924
1·45	0·4	0·379	0·379	0·652	0·656	0·979	
1·84	0·6	0·710	0·722	0·953	0·958		

Case $\rho_{12}=0{\cdot}8$		$n=12$		$n=25$		$n=100$	
$\gamma=\sigma_1/\sigma_2$	ρ_{XY}	Test (a)	Test (b)	Test (a)	Test (b)	Test (a)	Test (b)
1·00	0	0·100	0·100	0·100	0·100	0·100	0·100
1·06	0·1	0·116	0·115	0·138	0·137	0·262	0·260
1·13	0·2	0·167	—	0·254	0·250	0·642	0·643
1·21	0·3	0·254	—	0·437	0·436	0·922	0·926
1·30	0·4	0·379	0·375	0·652	0·660	0·979	—
1·55	0·6	0·710	0·735	0·953	0·964	—	—
2·08	0·8	0·965	0·981	1·000	1·000	—	—

Incomplete Beta Function. These values are compared in the table on p. 18 above with those for the likelihood ratio test. It will be seen that:

(1) When $\rho_{12} = 0$, i.e. when the two variables are known to be independent, the test based on ω or the ratio of sample variances is the better. This is a result already known, but the table shows how small is the difference between the tests.

(2) When ρ_{12} is 0·5 or 0·8, however, the likelihood ratio test is somewhat more sensitive for the small departures in γ from unity, and less sensitive for large departures than the ω test. Practically, this means that when ρ_{12} is known, the ω test is slightly the better, since we are most interested in situations where the chance of detection of a difference becomes large, say greater than 0·5 at any rate. It is possible that a test based on the criterion given in equation (25), if it could be obtained, would be more powerful than either of the other two tests when ρ_{12} is known. In practical cases, however, it will nearly always happen that ρ_{12} is unknown, and in such cases the r_{XY} of (18) appears the appropriate test criterion to use.

REFERENCES

Bose, S. (1935). *Sankhyā*, **1**, 65.

David, F. N. (1938). *Tables of the Correlation Coefficient. Biometrika* publication.

Finney, D. J. (1938). *Biometrika*, **30**, 190.

Fisher, R. A. (1915). *Biometrika*, **10**, 507.

Hirschfeld, H. O. (1937). *Biometrika*, **29**, 65.

Pearson, K. (1913). *Biometrika*, **9**, 1.

Neyman, J. & Pearson, E. S. (1936). *Statist. Res. Mem.* **1**, 1.

Tables of the Incomplete Beta Function (1934). *Biometrika* publication.

THE DISTRIBUTION OF RANGE IN SAMPLES FROM A NORMAL POPULATION, EXPRESSED IN TERMS OF AN INDEPENDENT ESTIMATE OF STANDARD DEVIATION

By D. NEWMAN

Department of Statistics, University College, London

1. Introduction

Starting from the contribution of Tippett (1925), a considerable amount of computational work has been carried out in recent years with the object of making possible the use of range, i.e. the distance between the highest and lowest observation, when dealing with samples from a normal population. Thus Tippett's tables of the mean range expressed in terms of the population standard deviation, σ, for sample sizes $n = 2$ to 1000 have been republished in *Tables for Statisticians and Biometricians*, Part II, Table XXII (K. Pearson, 1931). Later E. S. Pearson (1932) gave a table containing the standard deviation of range, and also the approximate upper and lower 10, 5, 1 and 0·5 % probability levels for sample sizes $n = 2$ to 100, again in terms of σ as unit. In doing this he used empirical Pearson-type curves with correct moments, and checked his results against some experimental sampling distributions.

If a number of small samples are available, it has been shown that a rapid estimate of σ may be obtained from the mean value of the range, which is only slightly less accurate than the estimate obtained from the sums of squares. Again, owing to the high correlation between range and standard deviation in a sample of size 10 or less, it was pointed out by Pearson & Haines (1935) that range may be usefully substituted for standard deviation in control charts used to study changes in the variation of quality in industry. In all these cases, however, the basic sampling distribution used has been that of the ratio of range to σ.

Not very long ago "Student" (the late Mr W. S. Gosset) suggested to Prof. E. S. Pearson that it might be useful to know more about the sampling distribution of the ratio

$$q = w/s,$$

where w is the range in a sample of n observations from a normal population with standard deviation σ, and s^2 is an *independent* and unbiassed estimate of σ^2 based on f degrees of freedom, obtained from a sum of squares in the usual manner. The type of problem which "Student" had in mind was one in which, as a result of an experiment, a number of 'treatment' means, say $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$, are available, and also an independent estimate, s^2, of their sampling variance. Then a rapid method of judging whether any treatment differences exist would consist in comparing the difference between the highest and lowest treatment means, say

$w = \bar{x}_n - \bar{x}_1$, with s. Should this difference be clearly significant, having regard to the values of n and f, the more divergent of the extremes, say $\bar{x}_1$, could be set aside, and the difference $\bar{x}_n - \bar{x}_2$ compared with s, using $n-1$ and f. This procedure would in fact be similar to that suggested by "Student" himself in his paper on "Errors in routine analysis" (1927, pp. 161–2), except that the ratio used would now be w/s rather than w/σ. Of course the probability levels for the former will tend to those of the latter ratio as $f \to \infty$.

In the following sections I shall first make use of the results obtained by Prof. Pearson in computing probability levels for w/σ, to determine appropriate levels for w/s, and then illustrate "Student's" suggestion on three practical examples. It should be noted that in a recent paper H. O. Hartley (1938) has suggested a systematic method of obtaining probability levels for "studentized" functions. It is hoped that before long fuller and more accurate tables may be available to supplement the present tables which rest to some extent on an empirical basis.

2. The expectation of $q = w/s$

Before describing the method of quadrature by which the probability levels were obtained, it may be useful to give a table from which the expectation of q for various values of n and f can be calculated. Since $s^2 = \sum\limits_{i=1}^{f+1}(x_i - \bar{x})^2/f$, we have for the probability distribution of s,

$$p(s) = \frac{f^{\frac{1}{2}f}}{2^{\frac{1}{2}(f-2)}\,\Gamma(\frac{1}{2}f)\,\sigma^f}\, s^{f-1} e^{-fs^2/2\sigma^2} \tag{1}$$

If we write $p(w)$ for the probability distribution of range, a function whose precise value is only known for the cases $n = 2, 3$, and denote an expectation by the symbol E, we have

$$\begin{aligned} E(q) &= \int_0^\infty p(q)\, q\, dq \\ &= \int_0^\infty \int_0^\infty w s^{-1} p(w)\, p(s)\, dw\, ds, \text{ since } w \text{ and } s \text{ are independent}, \\ &= \int_0^\infty w p(w)\, dw \times \int_0^\infty s^{-1} p(s)\, ds \\ &= E(w) \times \int_0^\infty s^{-1} p(s)\, ds. \end{aligned}$$

Since the values of $E(w/\sigma)$ for changing n have been tabled by Tippett, it is only necessary to consider the integral

$$\begin{aligned} \int_0^\infty \sigma s^{-1} p(s)\, ds &= \frac{f^{\frac{1}{2}f}}{2^{\frac{1}{2}(f-2)}\,\Gamma(\frac{1}{2}f)\,\sigma^{f-1}} \int_0^\infty s^{f-2} e^{-fs^2/2\sigma^2}\, ds \\ &= \sqrt{(\tfrac{1}{2}f)}\,\Gamma\{\tfrac{1}{2}(f-1)\}/\Gamma(\tfrac{1}{2}f). \end{aligned}$$

Hence it follows that

$$E(q) = E(w/\sigma) \times \sqrt{(\tfrac{1}{2}f)}\,\Gamma\{\tfrac{1}{2}(f-1)\}/\Gamma(\tfrac{1}{2}f). \tag{2}$$

A brief table of the second function is given below; the values of $E(w/\sigma)$ may be obtained from Tippett (1925), pp. 386–7, or from *Tables for Statisticians and Biometricians*, Part II, Table XXII.

TABLE I

Factors by which to multiply $E(w/\sigma)$ to obtain $E(q)$

Degrees of freedom	3	5	7	10	20	30	∞
Factor	1·382	1·189	1·126	1·084	1·040	1·026	1·000

3. COMPUTATION OF TABLE OF 5 AND 1% SIGNIFICANCE LEVELS FOR $q = w/s$

The problem is to determine, for different values of n and f, values q_α such that

$$\alpha = \int_{q_\alpha}^{\infty} p(q)\,dq$$
$$= \int_0^\infty \left\{p(w)\int_0^{w/q_\alpha} p(s)\,ds\right\} dw, \tag{3}$$

where $\alpha = 0{\cdot}05$ and $0{\cdot}01$. Since the distribution of q will clearly be independent of the population standard deviation σ, we may take σ as unity in the probability functions used in (3). Except in the cases $n = 2, 3$, which will be referred to again below, the procedure adopted was as follows:

(*a*) The ordinates of the empirical curves, say $y(w)$, obtained by E. S. Pearson (1932) for the cases $n = 4, 6, 10$, and 20 were used in place of the unknown $p(w)$. These ordinates had been calculated at equal intervals of 0·1 for w, the population standard deviation being the unit.

(*b*) Taking a trial value of q_α, the integrals $J(w, q_\alpha) = \int_0^{w/q_\alpha} p(s)\,ds$ were calculated, with the help of the *Tables of the Incomplete Gamma Function* (K. Pearson, 1922), for each value of w used in (*a*). Thus $J(w, q_\alpha) = I(u, p)$ in the notation of the tables, where

$$u = (\sqrt{\tfrac{1}{2}f})\left(\frac{w}{q_\alpha}\right)^2, \quad p = \tfrac{1}{2}f - 1. \tag{4}$$

(*c*) It was then necessary to apply quadrature to the products $y(w) \times J(w, q_\alpha)$, calculated at intervals 0·5 for w, through as much of the range $w = 0$ to $w = \infty$ as was necessary to obtain the required degree of accuracy.

(*d*) The resulting expression would of course not correspond to α exactly, as q_α had been guessed. Other trial values were taken for q_α, and the final value corresponding to $\alpha = 0{\cdot}05$ or $0{\cdot}01$ obtained by backward interpolation. Starting with the case $n = 2$, where the exact value of q_α could be obtained for all f, and knowing for $n > 2$ the limits to which q_α tended as $f \to \infty$, this process of trial and error was not found too laborious.

Special case when $n = 2$

In this case, taking $\sigma = 1$, the distribution of w assumes the simple form

$$p(w) = \frac{1}{\sqrt{\pi}} e^{-\frac{1}{4}w^2} \quad \text{for } 0 \leqslant w < \infty. \tag{5}$$

Hence the joint probability distribution of w and s is

$$p(w, s) = \frac{f^{\frac{1}{2}f}}{2^{\frac{1}{2}(f-2)}\pi^{\frac{1}{2}}\Gamma(\frac{1}{2}f)} s^{f-1} e^{-\frac{1}{4}w^2 - \frac{1}{2}fs^2} \tag{6}$$

Transforming to variables $q = w/s$ and s, since

$$\left| \frac{\partial(w, s)}{\partial(q, s)} \right| = s, \tag{7}$$

we obtain
$$p(q, s) = \text{constant} \times s^f e^{-\frac{1}{4}s^2(q^2+2f)}. \tag{8}$$

Now integrating for s between the limits 0 and ∞, we obtain for the probability distribution of q

$$p(q) = \sqrt{\left(\frac{2}{f\pi}\right)} \frac{\Gamma\{\frac{1}{2}(f+1)\}}{\Gamma(\frac{1}{2}f)} \left(1 + \frac{q^2}{2f}\right)^{-\frac{1}{2}(f+1)} \quad \text{for } q \geqslant 0. \tag{9}$$

This corresponds to the positive half of a "Student" distribution having f degrees of freedom. Values of q_α satisfying the relation $\alpha = \int_{q_\alpha}^{\infty} p(q)\,dq$ may therefore be obtained from R. A. Fisher's (1938) tables of the percentage points for t. Thus

$$q_\alpha = t_\alpha \sqrt{2}, \tag{10}$$

where t_α will be respectively the 5 and 1 % levels for t.*

Special case when $n = 3$

For this case McKay & Pearson (1933) have given an expression for the true distribution of w. The quadrature method employed when $n > 3$ was again used, but the true values of $p(w)$ were taken, and not the ordinates of the empirical curve upon which E. S. Pearson (1932) based his original tables of percentage limits for w.

The following table shows the framework of values for $q_{0 \cdot 05}$ and $q_{0 \cdot 01}$ obtained as has been described. Values for $f = \infty$ were of course already known, and values for $n = 2$ and 3 are exact.

From Table II, the more complete working Tables III and IV were obtained by interpolation. It was found that the changes in percentage levels, both with increasing n and f, ran most smoothly if the arguments $60/n$ and $60/f$ were used in place of n and f. On this basis, interpolation from the framework values was effected, using five and six-point Lagrangian formulae. Various checks were

* These levels correspond to deviations at which the ordinates cut off 2·5 and 0·5 % from each end of the t-distribution, but they are termed by Fisher the 5 and 1 % levels.

TABLE II

Framework values for $q_{0\cdot05}$ and $q_{0\cdot01}$

f \ n	5% points						1% points					
	2	3	4	6	10	20	2	3	4	6	10	20
5	3·64	4·60	5·22	6·03	7·00	8·21	5·70	6·99	7·83	8·94	10·26	11·95
10	3·15	3·88	4·34	4·92	5·60	6·47	4·48	5·27	5·77	6·42	7·21	8·22
20	2·95	3·58	3·97	4·45	5·01	5·71	4·02	4·63	5·01	5·50	6·08	6·82
30	2·89	3·49	3·86	4·31	4·82	5·47	3·89	4·45	4·78	5·23	5·75	6·40
∞	2·77	3·31	3·65	4·04	4·48	5·01	3·64	4·12	4·38	4·74	5·15	5·64

carried out, as for example that of comparing the values obtained by this method with the known true values in the case of $n = 2$. Finally, the figures were reduced to one place of decimals, and these are given in Tables III and IV.

4. Illustrative examples

In the following examples the range test is used as an alternative to the z test; the latter is, on theoretical grounds, the more efficient of the two in the sense that it is the more likely to detect the presence of real differences if they exist. Both "Student" and L. H. C. Tippett have, however, held that situations may be met, particularly when dealing with industrial problems, where the gain in speed following the use of range justifies the relatively small loss in efficiency. No doubt other types of examples besides those illustrated below will occur to the reader.

Example A

Fisher (1937, p. 93) has given the results of a 6×6 Latin square experiment in which six different fertilizer treatments were applied to a crop of potatoes. Denoting these treatments by the letters $A, B, \ldots, F$, the mean yields per plot in lb. were as follows:

A	B	C	D	E	F
345·0	426·5	477·8	405·2	520·2	601·8

The analysis of variance table is as follows:

Sources of variation	Degrees of freedom	Mean squares
Rows	5	10,839·72
Columns	5	4,893·45
Treatments	5	49,635·98
Error	20	1,527·05

TABLE III

5 % points for $q = w/s$

f \ n	2	3	4	5	6	7	8	9	10	11	12	20
5	3·64	4·6	5·2	5·7	6·0	6·3	6·6	6·8	7·0	7·2	7·3	8·2
6	3·46	4·4	4·9	5·3	5·6	5·9	6·1	6·3	6·5	6·7	6·8	7·6
7	3·34	4·2	4·7	5·1	5·4	5·6	5·8	6·0	6·2	6·3	6·5	7·2
8	3·26	4·1	4·5	4·9	5·2	5·4	5·6	5·8	5·9	6·0	6·2	6·9
9	3·20	4·0	4·4	4·7	5·0	5·2	5·4	5·6	5·7	5·8	5·9	6·6
10	3·15	3·9	4·3	4·7	4·9	5·1	5·3	5·5	5·6	5·7	5·8	6·5
11	3·11	3·8	4·3	4·6	4·8	5·0	5·2	5·4	5·5	5·6	5·7	6·3
12	3·08	3·8	4·2	4·6	4·8	5·0	5·1	5·3	5·4	5·5	5·6	6·2
13	3·06	3·7	4·2	4·5	4·7	4·9	5·0	5·2	5·3	5·4	5·5	6·1
14	3·03	3·7	4·1	4·4	4·6	4·8	5·0	5·2	5·3	5·4	5·5	6·0
15	3·01	3·7	4·1	4·4	4·6	4·8	4·9	5·1	5·2	5·3	5·4	6·0
16	3·00	3·6	4·1	4·4	4·6	4·7	4·9	5·0	5·1	5·2	5·3	5·9
17	2·98	3·6	4·0	4·3	4·5	4·7	4·8	5·0	5·1	5·2	5·3	5·8
18	2·97	3·6	4·0	4·3	4·5	4·7	4·8	5·0	5·1	5·2	5·3	5·8
19	2·96	3·6	4·0	4·3	4·5	4·6	4·8	4·9	5·0	5·1	5·2	5·8
20	2·95	3·6	4·0	4·2	4·5	4·6	4·7	4·9	5·0	5·1	5·2	5·7
24	2·92	3·5	3·9	4·2	4·4	4·6	4·7	4·8	4·9	5·0	5·1	5·6
30	2·89	3·5	3·9	4·1	4·3	4·5	4·6	4·7	4·8	4·9	5·0	5·5
40	2·86	3·4	3·8	4·0	4·2	4·3	4·5	4·6	4·7	4·8	4·9	5·4
60	2·83	3·4	3·8	4·0	4·2	4·3	4·4	4·5	4·6	4·7	4·8	5·2
∞	2·77	3·31	3·65	3·87	4·04	4·18	4·29	4·39	4·48	4·55	4·62	5·01

TABLE IV

1 % points for $q = w/s$

f \ n	2	3	4	5	6	7	8	9	10	11	12	20
5	5·70	7·0	7·8	8·5	8·9	9·3	9·6	10·0	10·3	10·5	10·7	12·0
6	5·24	6·4	7·0	7·5	7·9	8·3	8·5	8·9	9·1	9·3	9·5	10·5
7	4·95	5·9	6·5	7·0	7·4	7·7	7·9	8·2	8·4	8·6	8·7	9·6
8	4·74	5·7	6·2	6·6	6·9	7·2	7·4	7·7	7·9	8·1	8·2	9·0
9	4·60	5·4	6·0	6·3	6·6	6·9	7·1	7·3	7·5	7·7	7·8	8·6
10	4·48	5·3	5·8	6·1	6·4	6·7	6·8	7·0	7·2	7·4	7·5	8·2
11	4·39	5·1	5·6	5·9	6·2	6·4	6·6	6·8	7·0	7·2	7·3	8·0
12	4·32	5·0	5·5	5·8	6·1	6·3	6·5	6·7	6·8	6·9	7·0	7·7
13	4·26	5·0	5·4	5·7	6·0	6·2	6·4	6·6	6·7	6·8	6·9	7·6
14	4·21	4·9	5·3	5·7	5·9	6·1	6·2	6·4	6·5	6·6	6·7	7·4
15	4·17	4·8	5·2	5·6	5·8	6·0	6·1	6·3	6·4	6·5	6·6	7·3
16	4·13	4·8	5·2	5·5	5·7	5·9	6·0	6·2	6·3	6·4	6·5	7·2
17	4·10	4·7	5·1	5·5	5·7	5·9	6·0	6·2	6·3	6·4	6·5	7·1
18	4·07	4·7	5·1	5·4	5·6	5·8	5·9	6·1	6·2	6·3	6·4	7·0
19	4·05	4·7	5·0	5·3	5·5	5·7	5·8	6·0	6·1	6·2	6·3	6·9
20	4·02	4·6	5·0	5·3	5·5	5·7	5·8	6·0	6·1	6·2	6·3	6·8
24	3·96	4·5	4·9	5·2	5·4	5·6	5·7	5·8	5·9	6·0	6·1	6·6
30	3·89	4·5	4·8	5·0	5·2	5·4	5·5	5·6	5·8	5·8	5·9	6·4
40	3·82	4·4	4·7	4·9	5·1	5·3	5·4	5·5	5·6	5·7	5·8	6·2
60	3·76	4·3	4·6	4·8	5·0	5·1	5·2	5·3	5·4	5·5	5·6	6·0
∞	3·64	4·12	4·38	4·59	4·74	4·87	4·98	5·07	5·15	5·22	5·28	5·64

To test the significance of treatment differences as a whole we find $z = 1{\cdot}7407$, while for degrees of freedom $f_1 = 5, f_2 = 20$, R. A. Fisher's tables give $z_{0{\cdot}01} = 0{\cdot}7058$. There are clearly, therefore, significant treatment differences present. The individual treatment means have been plotted in the accompanying figure, and the tabled significance levels for $q = w/s$ may be used, with discretion, as a test-rule in investigating the situation.

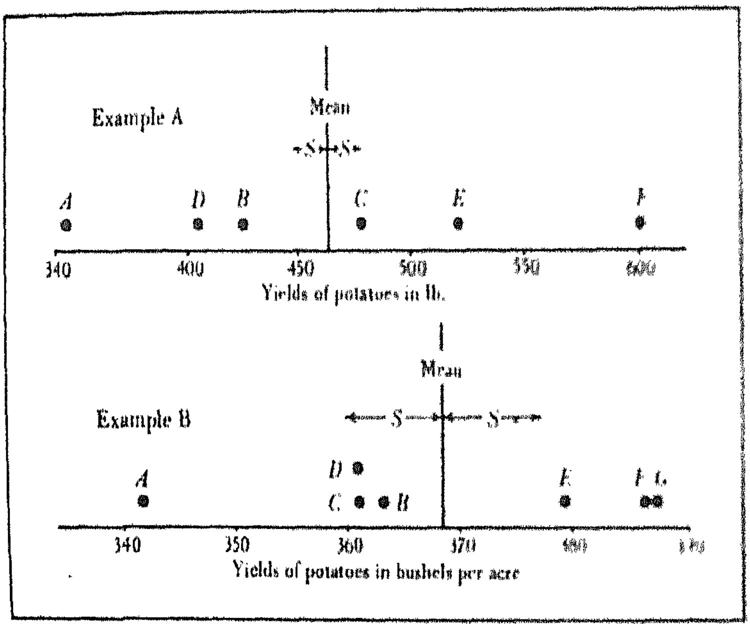


Fig. 1.

In the first place the appropriate pooled estimate, based on 20 degrees of freedom, of the standard error of a treatment mean is

$$s = \frac{39{\cdot}08}{\sqrt{6}} = 15{\cdot}95.$$

The range of the six treatment means is

$$w = 601{\cdot}8 - 345{\cdot}0 = 256{\cdot}8,$$

so that
$$q = \frac{w}{s} = 16{\cdot}1.$$

Table IV above gives for $n = 6, f = 20$ a 1 % level for q of 5·5, confirming the very significant scatter of treatment means brought out by the z-test.

We may now ask whether, if we were to exclude the most divergent treatment F, there is evidence of a significant difference among the remaining five treatments. We find that $w = 520{\cdot}2 - 345{\cdot}0 = 175{\cdot}2$, and, using the same estimate of standard

error, $s = 15\cdot95$, find that $q = 11\cdot0$. This value is still well beyond $q_{0\cdot01} = 5\cdot3$, the figure obtained from Table IV with $n = 5, f = 20$.

Making successive trials we find that:

(i) Omitting A and F, the range of the four treatments B, C, D, and E is still significant, since $q = 115\cdot0/15\cdot95 = 7\cdot2$, while for $n = 4, f = 20$ we have $q_{0\cdot01} = 5\cdot0$.

(ii) Omitting A, E, and F, the range of the three treatments B, C, and D is significant at the 5 % but just not significant at the 1 % level. For in this case $q = 72\cdot6/15\cdot95 = 4\cdot5$, while for $n = 3, f = 20$ we find $q_{0\cdot05} = 3\cdot6$ and $q_{0\cdot01} = 4\cdot6$.

(iii) On the other hand, if we divided the six treatments into two groups, one consisting of A, D, B, and the other of C, E, F, we find that the value of q in both groups falls beyond $q_{0\cdot01} = 4\cdot6$.

We are therefore led to conclude that the high value of z obtained from the comprehensive test cannot be explained by one, or even two, treatments differing from the others. It is doubtful, even, whether any three treatments out of the six could be regarded as forming a homogeneous group.

Two final points should be noted. In the first place, after omitting successive treatments regarded as divergent, analysis of variance procedure could be applied to test for significant differences among the remaining treatments. The calculation would, however, not be as quick as that involved in the successive trials (i), (ii), and (iii) above, using q. Finally, as mentioned above, the method followed, whether z or q is used, must be employed with discretion, as is always the case when observations are rejected and a hypothesis tested using the selected data that remain.

Example B

A similar example has been taken from Snedecor (1937, p. 214). The following are seven treatment means expressed in terms of bushels per acre, obtained from a 7×7 Latin square experiment, again comparing the effect of different fertilizers on potatoes.

A	*B*	*C*	*D*	*E*	*F*	*G*
341·86	363·14	360·57	360·43	379·86	386·29	387·14

The appropriate estimate of the standard deviation of a mean of seven plots calculated from the error sum of squares of the analysis of variance table is $s = 9\cdot52$. Testing the significance of treatment differences as a whole, it is found that $z = 0\cdot5574$, while for degrees of freedom $f_1 = 6, f_2 = 30$, $z_{0\cdot05} = 0\cdot4420$ and $z_{0\cdot01} = 0\cdot6226$. The ratio $q = w/s$ for all seven treatments is $45\cdot3/9\cdot52 = 4\cdot7$ a value lying between $q_{0\cdot05} = 4\cdot5$ and $q_{0\cdot01} = 5\cdot4$ (entering Tables III and IV with $n = 7, f = 30$). Thus using either test we should conclude that there were probably significant treatment differences.

The seven treatment means have been plotted in the lower half of the figure. There is a suggestion that either if (i) treatments F and G or (ii) treatment A

were regarded as exceptional, the remaining treatments would form a homogeneous group.

This is confirmed on investigation:

(i) Removing F and G, we find $q = 38{\cdot}0/9{\cdot}52 = 4{\cdot}0$, which is just below the 5 % level ($q_{0\cdot05} = 4{\cdot}1$) for $n = 5, f = 3$.

(ii) Removing A, we find $q = 26{\cdot}7/9{\cdot}52 = 2{\cdot}8$, which is well below the 5 % level ($q_{0\cdot05} = 4{\cdot}3$) for $n = 6, f = 3$. The more homogeneous group appears to be left in the second case, i.e. on removing A alone. Beyond this indication we cannot go, as it would need a knowledge of the character of the treatments to draw more definite conclusions.

Example C

In cases where a number of duplicate observations are available, an estimate of variability may be obtained rapidly from summing the squares of the differences between pairs. Thus

$$s^2 = \sum_{t=1}^{k} (x_{t1} - x_{t2})^2/2k$$

and has k degrees of freedom. We may now compare the range in the means of pairs with s, in order to determine whether there is too much variation between pairs having regard to the variation within pairs. It may be noted that an estimate of σ may be obtained even more rapidly by calculating the mean range in pairs and multiplying by 0·8862,* so that

$$s' = 0{\cdot}8862 \times \sum_{t=1}^{k} | x_{t1} - x_{t2} |/k,$$

but since the sampling distribution of $(s')^2$ is not that of χ^2, we cannot justifiably take $q = w/s'$ and refer to the tables of percentage limits I have given.

Determinations of percentage fibre

Analyst	A	B	C	D	E	F	G	H	I	J
1st determination (x_{t1})	12·66	12·51	12·32	13·15	12·73	12·48	12·30	12·14	12·48	13·30
2nd determination (x_{t2})	12·47	12·62	12·55	12·93	12·43	12·46	12·73	12·03	12·44	12·68
Difference (d_t)	0·19	−0·11	−0·23	0·22	0·30	0·02	−0·43	0·11	0·04	0·62
Sum	25·13	25·13	24·87	26·08	25·16	24·94	25·03	24·17	24·92	25·98

* The reciprocal of 1·12838, since the expectation of range in a sample of two individuals is 1·12838σ.

The data shown on p. 28 above consist of ten duplicate determinations of the percentage fibre in carefully mixed samples taken from the same supply of Soya Cotton Cake, each pair of values being obtained by one analyst.* Ten different analysts were concerned. The problem is to determine whether these few observations provide any evidence of systematic differences in technique between the analysts.

The full analysis of variance is as follows:

	Sum of squares	D.F.	Mean square
Within pairs	0·411450	10	0·041145
Between pairs	1·352045	9	0·150227
Total	1·763495	19	

Testing for differences between analysts, we obtain $z = 0{\cdot}6475$, a value falling between the 1 % (0·7990) and 5 % (0·5527) levels.

Using the range method, we note that the estimate of the variance of the *sum* of two determinations is

$$\frac{1}{k}\sum_{i=1}^{k}(x_{i1}-x_{i2})^2 = \frac{1}{10}\sum_{i=1}^{10}(d_i^2) = 0{\cdot}08229,$$

giving an estimate of the standard error of a sum of 0·2868.† The range in the ten sums in the table is $w = 26{\cdot}08 - 24{\cdot}17 = 1{\cdot}91$, so that $q = 1{\cdot}91/0{\cdot}2868 = 6{\cdot}66$. For $n = 10$, $f = 10$, Tables III and IV show $q_{0{\cdot}05} = 5{\cdot}6$, $q_{0{\cdot}01} = 7{\cdot}2$. Thus the difference is significant at the 5 % level, a result similar to that found using the z-test.

If now we omit determinations of analyst D (who gave the highest readings) we find $q = (25{\cdot}98 - 24{\cdot}17)/0{\cdot}2868 = 6{\cdot}31$, and is still significant at the 5 % level, since for $n = 9$, $f = 10$, $q_{0{\cdot}05} = 5{\cdot}5$. If, however, we omit the determinations of analyst H (who gave the lowest readings) we find $q = (26{\cdot}08 - 24{\cdot}87)/0{\cdot}2868 = 4{\cdot}22$, and is no longer significant. We should conclude, therefore, that except possibly in the case of analyst H, there is no evidence on the data available of systematic differences in technique.

I should like to take this opportunity of thanking Prof. E. S. Pearson not only for suggesting the problem to me, but also for his kind help in putting this paper together.

* The figures have been taken from more extensive data made available through the kindness of Dr J. F. Tocher.

† The corresponding estimate obtained, as described on p. 28 above, from the mean difference between pairs is 0·291.

REFERENCES

FISHER, R. A. (1938). *Statistical Methods for Research Workers*. 7th ed. Edinburgh: Oliver and Boyd.

—— (1937). *The Design of Experiments*, 2nd ed. Edinburgh: Oliver and Boyd

HARTLEY, H. O. (1938). *J.R. Statist. Soc. Suppl.* **5**, 80.

McKAY, A. T. & PEARSON, E. S. (1933). *Biometrika*, **25**, 415.

PEARSON, E. S. (1926). *Biometrika*, **18**, 173.

—— (1932). *Biometrika*, **24**, 404.

PEARSON, E. S. & HAINES, JOAN (1935). *J.R. Statist. Soc. Suppl.* **2**, 83.

PEARSON, KARL (1922). *Tables of the Incomplete Gamma Function*. London: Biometrika Office.

—— (1931). *Tables for Statisticians and Biometricians*, Part II. London: Biometrika Office.

SNEDECOR, G. W. (1937). *Statistical Methods*. Iowa Collegiate Press, Inc.

"STUDENT" (1927). *Biometrika*, **19**, 151.

TIPPETT, L. H. C. (1925). *Biometrika*, **17**, 364.

ON A COMPREHENSIVE TEST FOR THE HOMOGENEITY OF VARIANCES AND COVARIANCES IN MULTIVARIATE PROBLEMS

By D. J. BISHOP

Department of Statistics, University College, London

1. Introduction

Now that satisfactory and probably final solutions have been obtained for a wide variety of statistical problems concerned with a single normally distributed variable, more and more attention has recently been given to the solution of multivariate problems. The multiple correlation methods of the old large sample theory have been replaced in many instances by others for which "studentized" test criteria are available, often having sampling distributions that are already familiar in univariate problems. In a recent paper on "The statistical utilization of multiple measurements", R. A. Fisher (1938*a*) has shown the connexion between certain of these methods: the D^2-statistic work of Mahalanobis, the discriminant function methods of the Galton Laboratory and the generalized "Student's" ratio of Hotelling. A similar very general problem was dealt with some time ago by S. S. Wilks (1932), while mention may also be made of two papers by D. G. Lawley (1938*a*, *b*) and a paper by P. L. Hsu (1938). The purpose of the methods put forward is to obtain information regarding the mean values of a number, say q, of correlated variables in one or more, say k, populations from which random samples have been drawn. If we denote by x_s a value of the sth variable ($s = 1, 2, \ldots, q$), then in all this work it has been assumed not only that x_s is normally distributed, but that it has the same variance σ_s^2 in every population sampled. Further, it is assumed that if x_u is a second variable the correlation coefficients ρ_{su} between x_s and x_u is the same in all populations. The estimates of variance and covariance required in order to "studentize" the function of the sample means are therefore obtained by pooling together the sums of squares and sums of products from all samples. While it is true that even if σ_s and ρ_{su} are not the same in all populations the error involved may not be very large, it is however important to have available some means of testing the basic hypothesis which assumes homogeneity throughout the populations.

Such a test has been derived by S. S. Wilks (1932) by an extension of Neyman & Pearson's likelihood ratio method of approach. Hitherto the somewhat lengthy computations required to obtain the moments of the sampling distribution of the test criterion have probably discouraged its use. The objects of the present paper are as follows:

(*a*) In the simple but commonly met case, where the k samples are of the same

size, to put Wilks' test into such a form that once the lengthy process of computing the kq variances and $\frac{1}{2}kq(q-1)$ covariances has been carried out, relatively little further labour is required to obtain a test criterion which may be referred with practical accuracy either to Fisher's z-tables or to the *Tables of the Incomplete Beta Function* (K. Pearson, 1934).

(*b*) In the case where the sample sizes differ, to suggest an alternative procedure which is accurate when dealing with large samples.

It is of course always open to question how far a single comprehensive criterion is satisfactory in a complex problem. Certain points should, however, be remembered. In the problem referred to above, dealing with the means only, the usefulness of a single criterion has been widely recognized. If, when applied to adequately large samples, a suitably chosen comprehensive criterion shows no significant evidence of lack of homogeneity among the variances and covariances, we are saved the lengthy process of making many individual comparisons. If, however, the criterion falls beyond the significance level, it will be necessary, as when dealing with the means, to make a more detailed analysis in order to locate the source of disturbance.

Finally, it may be noted that in the case of a single variate ($q = 1$) the problem is much simplified. A full discussion of two tests available in this case, the Neyman & Pearson L_1 test and the Bartlett μ test has recently been published elsewhere (Bishop & Nair, 1939).

2. Wilks' generalized likelihood criterion

It is proposed first to define this criterion in the simple case when the samples from each of the k populations are of the same size, n; to quote the sampling moments derived by Wilks; to give the form of a working test developed in the later pages; and to illustrate its use on an example.

It will be convenient to use the following notation. We shall consider k samples, each of size n, drawn from populations of q variates. That is to say, the total number of individuals sampled is $kn = N$ and for each individual q characters are measured, so that the total number of observations is Nq. Let x_{sti} be the observed value of the character s of the ith individual in the tth sample. Then $s = 1, 2, \ldots, q$; $t = 1, 2, \ldots, k$; and $i = 1, 2, \ldots, n$.

Let
$$x_{st.} = \frac{1}{n}\sum_{i=1}^{n} x_{sti}$$

be the mean value observed for the character s in the tth sample. Now put

$$v_{sut} = \frac{1}{n}\sum_{i=1}^{n}(x_{sti}-x_{st.})(x_{uti}-x_{ut.}),* \qquad \ldots\ldots(1)$$

* The use of the divisor n instead of the more usual $n-1$ avoids the necessity for a multiplying constant in the expression for λ_1.

where $s, u = 1, 2, \ldots, q$. Then if $s = u$, v_{sut} is just the variance of observations of character $s = u$ within the tth sample, whereas if $s \neq u$, v_{sut} is the covariance of the characters s and u.

The generalized variance in the tth sample is defined by the symmetrical determinant

$$|v_{sut}| = \begin{vmatrix} v_{11t}, & v_{12t}, & \ldots, & v_{1qt} \\ v_{12t}, & v_{22t}, & \ldots, & v_{2qt} \\ \vdots & \vdots & & \vdots \\ v_{1qt}, & v_{2qt}, & \ldots, & v_{qqt} \end{vmatrix}. \qquad \ldots\ldots(2)$$

Then if

$$v_{su} = \frac{1}{k}\sum_{t=1}^{k} v_{sut}, \qquad \ldots\ldots(3)$$

the likelihood criterion appropriate for testing the comprehensive hypothesis, say H_0, that all the corresponding variances and covariances in the k sampled populations are equal is

$$\lambda_1 = \frac{\prod_{t=1}^{k} |v_{sut}|^{\frac{1}{2}n}}{|v_{su}|^{\frac{1}{2}N}}. \qquad \ldots\ldots(4)$$

It has been found convenient to study the sampling distribution of some fractional power of λ_1, rather than that of λ_1 itself, owing to the extreme skewness of the latter distribution. Within limits there is some choice in the power of λ_1 which may be selected to give a convenient sampling distribution. Throughout this paper we shall follow Pearson & Wilks (1933) by using the $1/N$th power, so that the criterion for use in practice will be

$$\lambda_1^{1/N} = l_1 = \sqrt{\frac{\prod_{t=1}^{k} |v_{sut}|^{1/k}}{|v_{su}|}}. \qquad \ldots\ldots(5^*)$$

If H_0, the hypothesis tested, be true, the hth moment about zero of the sampling distribution of l_1 has been given by Wilks (1932, p. 490) in the form

$$\mu_h' = k^{\frac{1}{2}qh}\left\{\frac{(\Gamma(\frac{1}{2}(n+h/k-s)))^k\,\Gamma(\frac{1}{2}(k(n-1)+1-s))}{(\Gamma(\frac{1}{2}(n-s)))^k\,\Gamma(\frac{1}{2}(k(n-1)+1+h-s))}\right\}. \qquad \ldots\ldots(6)$$

As a result of trial in a number of cases it seems probable that the Pearson Type I distribution in the form

$$p(l_1) = \frac{\Gamma(m_1+m_2)}{\Gamma(m_1)\,\Gamma(m_2)} l_1^{m_1-1}(1-l_1)^{m_2-1}, \text{ for } 0 \leqslant l_1 \leqslant 1, \qquad \ldots\ldots(7)$$

will give a good approximation to the distribution of l_1, if the parameters m_1 and

* Note that $\{l_1(q=1)\}^2 = L_1$, in the original notation of Neyman & Pearson (1931).

m_2 are so chosen that the first two moments of (7) agree with the true values as given by (6). If this is done it is found that

$$m_1 = \frac{\mu_1'(\mu_1' - \mu_2')}{\mu_2}, \qquad \text{......(8)}$$

and

$$m_2 = \frac{(1-\mu_1')(\mu_1' - \mu_2')}{\mu_2}, \qquad \text{......(9)}$$

where

$$\mu_2 = \mu_2' - (\mu_1')^2, \qquad \text{......(10)}$$

μ_1' and μ_2' being respectively the first and second moments about zero of l_1. From the relation (6) it is seen that their values are

$$\mu_1' = k^{kq} \prod_{s=1}^{q} \left\{ \frac{(\Gamma(\tfrac{1}{2}(n+1/k-s)))^k \, \Gamma(\tfrac{1}{2}(k(n-1)+1-s))}{(\Gamma(\tfrac{1}{2}(n-s)))^k \, \Gamma(\tfrac{1}{2}(k(n-1)+2-s))} \right\} \qquad \text{......(11)}$$

and

$$\mu_2' = k^{q} \prod_{s=1}^{q} \left\{ \frac{(\Gamma(\tfrac{1}{2}(n+2/k-s)))^k \, \Gamma(\tfrac{1}{2}(k(n-1)+1-s))}{(\Gamma(\tfrac{1}{2}(n-s)))^k \, \Gamma(\tfrac{1}{2}(k(n-1)+3-s))} \right\}. \qquad \text{......(12)}$$

The hypothesis H_0 will be rejected when l_1 is exceptionally low. The probability level may be obtained from the *Tables of the Incomplete Beta Function*. Alternatively 5 % and 1 % levels for l_1 can be obtained from R. A. Fisher's z-tables (1938*b*, Table VI), by writing

$$l_1 = \frac{f_2}{f_2 + f_1 e^{2z}}, \qquad \text{......(13)}$$

where z has degrees of freedom

$$f_1 = 2m_2, \quad f_2 = 2m_1.* \qquad \text{......(14)}$$

It is no doubt the labour involved in calculating the moments μ_1' and μ_2' that has discouraged the use of Wilks' test. As will be shown in the later sections of the paper, if n is not too small,† the following empirical relations may be used with sufficient accuracy for practical purposes to express the m_1 and m_2 of (14) directly in terms of (*a*) the number of variables q, (*b*) the sample size n and (*c*) the number of samples k:

$$m_1 = k(n-q) - 0{\cdot}01(k-1)(90 - 39q + 9q^2), \qquad \text{......(15)}$$

$$m_2 = 0{\cdot}25(k-1)\,q(q+1). \qquad \text{......(16)}$$

These formulae are found to give satisfactory results for all values of k, over the range considered for q, namely 1 to 6.

In the following sections of the paper we shall first give an example illustrating the practical use of the test and afterwards justify the approximations which have been made.

* Since in general f_1 and f_2 in (14) will not be integral, it is necessary to interpolate in the tables for fractional degrees of freedom.

† Certainly if $n \geqslant 20$, and probably if $n \geqslant 10$ provided $q < 5$.

3. Illustrative example

The data used were taken from the appendix of a craniological study by E. Pittard (1909) and consist of measurements of skulls of males found in different localities in the Rhone valley. The characters considered are:

(1) Length of skull in mm., denoted by L.
(2) Breadth of skull in mm., denoted by B.
(3) Height of skull in mm., denoted by H.
(4) Breadth of face in mm., denoted by F.

The skulls are divided into five groups of thirty skulls each, according to the neighbourhood in which they were found. The five groups are:

(1) Skulls found in the village of Biel.
(2) ,, ,, ,, ,, Naters.
(3) ,, ,, ,, ,, Viège.
(4) ,, ,, ,, ,, Rarogne.
(5) ,, ,, ,, ,, Sierre.

It is desired to investigate if the population variances and covariances of the measurements of the four characters may be taken as the same for each of the five populations sampled. This is equivalent to asking whether the population variances and covariances differ from village to village.

With the notation previously employed it is clear that in this case $n = 30$, $k = 5$ and $q = 4$. The sample variances and covariances, as given by (1), were calculated in the usual manner. Their values, together with the sample correlation coefficients, are shown in Table IA. Using these results, the values of the generalized sample variances $|v_{sut}|$, which are also given in Table IA, were calculated from equation (2). The substitution of these values of $|v_{sut}|$ into equation (5) gives $l_1 = 0{\cdot}725$.

The probability levels of the sampling distribution of l_1 are given by (13) and (14), where

$$m_1 = 126{\cdot}880 \quad \text{and} \quad m_2 = 20{\cdot}000$$

are determined by substituting $k = 5$, $n = 30$ and $q = 4$ in the empirical relations (15) and (16). The levels are found to be:

$$5\ \%\text{ level of } z = 0{\cdot}1830 \text{ and } 1\ \%\text{ level of } z = 0{\cdot}2573,$$

hence $$5\ \%\text{ level of } l_1 = 0{\cdot}815 \text{ and } 1\ \%\text{ level of } l_1 = 0{\cdot}791.$$

The calculated value of l_1 being considerably below the 1 % level, the hypothesis that the variances and covariances are the same for each of the five populations sampled must be rejected.

Further tests may now be made in order to find where the lack of uniformity

TABLE IA

s	u	v_{su1}	v_{su2}	v_{su3}	v_{su4}	v_{su5}	v_{su6}
L	L	30·2222	46·1333	46·9123	29·8267	34·0000	39·2387
B	B	25·4766	28·4989	22·5789	43·4622	21·7556	28·3544
H	H	20·6989	18·3733	23·0156	41·1289	32·4933	29·1229
F	F	32·6489	25·8767	31·4322	25·4900	10·6722	26·4240
L	B	22·0333	1·5000	10·5212	−2·4467	10·6444	8·4544
L	H	1·7444	8·9000	0·3689	19·4467	0·2000	6·1440
L	F	28·0556	7·2333	12·5922	0·3867	−0·3278	9·5800
B	H	1·8433	10·5867	−2·2689	16·0822	−7·2000	3·8987
B	F	20·7400	12·4567	7·3744	17·5067	3·2111	12·2578
H	F	8·9844	1·6467	−10·0004	15·3300	0·9000	3·3721
		$\lvert v_{su1} \rvert$	$\lvert v_{su2} \rvert$	$\lvert v_{su3} \rvert$	$\lvert v_{su4} \rvert$	$\lvert v_{su5} \rvert$	$\lvert v_{su6} \rvert$
		138,521	307,934	510,165	422,517	296,115	585,487
s	u	r_{su1}	r_{su2}	r_{su3}	r_{su4}	r_{su5}	
L	B	0·69702	0·04137	0·32327	−0·06706	0·39981	
L	H	0·05111	0·30570	0·01101	0·55522	0·00781	
L	F	0·78401	0·20935	0·32702	0·01402	−0·01375	
B	H	0·06701	0·46265	−0·09764	0·38038	−0·27080	
B	F	0·71913	0·45871	0·27681	0·52597	0·16861	
H	F	0·28853	0·07552	−0·36475	0·47346	0·03867	

occurs. To decide whether the variances differ we may apply four single variate L_1 tests, treating the observations of each character separately. When this is done the results following in Table IB are obtained.

The tables of P. P. N. Nayer (1936) give the 5 % level for L_1, when $n = 30$ and $k = 5$, as 0·936, so that there is no evidence to suggest that the variances of measurements of any one character differ significantly from sample to sample.

The lack of uniformity must therefore occur in the correlations, so that it is necessary to examine the variation in each of the six sets of five correlation

TABLE IB

Character	L_1
L	0·985
B	0·967
H	0·964
F	0·974

coefficients. If r be a sample correlation coefficient calculated from n pairs of observations it is known that

$$z' = \tfrac{1}{2}(\log_e(1+r) - \log_e(1-r))$$

is approximately normally distributed with standard deviation $1/\sqrt{(n-3)}$, the mean of z' being a function of the population correlation coefficient and the sample size n. Consequently, we may test whether correlation coefficients r_{sut} $(t = 1, 2, \ldots, k)$, each based on n pairs of observations and obtained from k independent samples, differ only through chance fluctuations from some common population value ρ_{su}, by calculating

$$\chi^2 = (n-3) \sum_{t=1}^{k} (z'_t - \bar{z}')^2,$$

where
$$\bar{z}' = \frac{1}{k} \sum_{t=1}^{k} z'_t,$$

and entering the tables of the χ^2 integral with $k-1$ degrees of freedom. A significantly large value of χ^2 will indicate that the k samples cannot be considered as having been drawn from populations with a common value of ρ. This procedure was carried out with the results shown in Table Ic.

TABLE Ic

Correlation between characters	χ^2	Remarks
L and B	14·31	Significant
L ,, H	7·82	
L ,, F	20·47	Significant
B ,, H	11·67	Probably significant
B ,, F	8·81	
H ,, F	12·08	Probably significant

The probability levels for the distribution of χ^2 with $f = 4$ are: 5 % level = 9·49, 2 % level = 11·67 and 1 % level = 13·28. It is seen that all the six calculated values of χ^2 are above the expectation value of 4. There is significant variation in the correlation coefficients r_{LBt} and r_{LFt}, whilst it is largely a matter of personal opinion as to whether the suggestion of lack of uniformity in r_{BHt} and r_{HF} shall be judged significant or not.

To summarize: analysis of the data leads to the following conclusions:

(*a*) The comprehensive test shows evidence of significant variation from sample to sample of the variances and covariances.

(*b*) This lack of uniformity is not due to differences in the population variances.

(*c*) There is clear evidence of heterogeneity among some of the correlation coefficients, in particular for those between length and breadth of skull and between length of skull and breadth of face. Results of this kind are frequently met with when dealing with craniological data, the correlation coefficient often being subject to a considerable degree of instability. The meaning to be attached to these fluctuations in the correlations, when considered from a craniological viewpoint, seems to be somewhat obscure.

The above example may give the impression that a great deal of labour is required, even after the comprehensive test is used. If, however, the comprehensive test had provided no evidence for rejecting the hypothesis H_0, it would, of course, not have been necessary to make the four single-variate L_1 tests and the six correlation tests. Even when the hypothesis H_0 is rejected and it is necessary to apply separate tests to find the causes of the lack of uniformity, the only labour which will have been wasted is the relatively small amount involved in the calculation of the determinants $|r_{st}|$. The really lengthy computation is that required to obtain the sums of squares and sums of products on which Table I A is based; this cannot be avoided if a detailed analysis is desired.

4. The adequacy of the Type I approximation

The work now falls into two stages, which will be concerned with

(*a*) the adequacy of the Type I curve of equation (7) to represent the unknown true distribution of l_1;

(*b*) the accuracy of the empirical formulae (15) and (16) for m_1 and m_2.

The hope that the Type I form of curve might give an adequate approximation was based on two main considerations:

(*a*) Values of l_1 are restricted and can only lie between zero and unity.

(*b*) On other occasions the use of the principle of likelihood has yielded criteria which are either exactly distributed in the Type I form, or are so distributed that a good approximation has been obtainable by the use of this kind of curve.

In order to compare the true distribution with the approximate form, the first four moments of both distributions have been calculated in a number of cases. The manner of choosing the approximate distribution ensures that its first two moments will agree with those of the true distribution, and an idea of the accuracy of the approximation may be gathered by comparing the third and fourth moments of the true distribution with the corresponding moments of the approximate form. The distribution is such that a comparison of moments calculated about the mean is easier than that obtained when moments taken about zero are used. The cases considered fall into two groups:

(i) $k = 5$; $n = 10$, 20, 40 and 50; $q = 1$, 2, 3, 4, 5 and 6

and (ii) $n = 30$; $k = 2$, 5 and 10; $q = 1$, 2, 3, 4, 5 and 6.

The process of calculation was as follows:

(*a*) The first four moments about zero of the true distribution were obtained by using 10-figure tables of logarithms of the Gamma function (E. S. Pearson, 1922). A large number of figures were required in the values of these moments, since the dispersion of many of the distributions was small.

(*b*) The values of μ_2, μ_3 and μ_4, the second, third and fourth moments about the mean, were then obtained in the usual way and the constants m_1 and m_2 were calculated by using (8) and (9).

(*c*) The third and fourth moments about zero of the Type I distribution (7) are given by

$$ {}_{\mathrm{I}}\mu_3' = \frac{m_1(m_1+1)(m_1+2)}{(m_1+m_2)(m_1+m_2+1)(m_1+m_2+2)} = \frac{(m_1+2)}{(m_1+m_2+2)}\mu_2', \quad \ldots\ldots(17) $$

and

$$ {}_{\mathrm{I}}\mu_4' = \frac{m_1(m_1+1)(m_1+2)(m_1+3)}{(m_1+m_2)(m_1+m_2+1)(m_1+m_2+2)(m_1+m_2+3)} = \frac{(m_1+3)}{(m_1+m_2+3)}\,{}_{\mathrm{I}}\mu_3' \quad \ldots\ldots(18) $$

so that, using values of m_1 and m_2 already obtained, these may be calculated.

(*d*) From the values of ${}_{\mathrm{I}}\mu_3'$ and ${}_{\mathrm{I}}\mu_4'$ the corresponding moments about the mean, ${}_{\mathrm{I}}\mu_3$ and ${}_{\mathrm{I}}\mu_4$, were calculated. Comparing μ_3 and μ_4 with ${}_{\mathrm{I}}\mu_3$ and ${}_{\mathrm{I}}\mu_4$, respectively, gives an idea of the adequacy of the approximation.

Table II gives the results obtained; in addition, values of $\beta_1 = \mu_3^2/\mu_2^3$ and $\beta_2 = \mu_4/\mu_2^2$ are included as measures of the skewness and kurtosis of the distributions with which we are dealing.

It is seen from these tables that the distributions represented vary widely in shape, but as far as can be judged by comparison of third and fourth moments, the use of the Type I approximation would seem to be quite satisfactory in every case considered. The cases dealt with cover a fairly wide field and have not been chosen with a view to obtaining especially good agreement of the moments.

The inclusion of curves for which $q = 1$ may be questioned having regard to the fact, already stated, that the single variate problem has been fully treated. We must therefore remark that, as $L_1 = \{l_1(q=1)\}^2$, the fact that the distribution of L_1 may be represented by a Type I curve does not imply that this kind of approximation will work adequately for the distribution of $l_1 = L_1^{\frac{1}{2}}$. Moreover, some of the distributions obtained by putting $q = 1$ are abnormally skew and highly leptokurtic. As it happens, when $q = 1$ the Type I curve appears to give as good an agreement as for other values, but the exclusion of this set of distributions might have had the effect of making the approximation appear better than was really the case, in that it would have removed a number of the "more difficult" distributions. Thus, although the criterion L_1 will probably always be used in preference to $l_1(q=1) = L_1^{\frac{1}{2}}$, it seemed desirable to include distributions

TABLE II

Moment coefficients of l_1 in trial cases

k = no. of populations sampled, q = no. of variables, n = size of samples

q		$k=5$, $n=10$	$k=5$, $n=20$	$k=5$, $n=30$	$k=5$, $n=40$	$k=5$, $n=50$
1	μ_1'	0·9551,5592	0·9788,4625	0·9861,620246	0·9807,18[illegible]	0·9818,298271
	μ_2	0·0009,3729	0·0002,1661	0·0000,937517	0·0000,52[illegible]	0·0000,330404
	μ_3	−0·0000,3646	−0·0000,0429	−0·0000,012438	−0·0000,005188	−0·0000,002638
	${}_1\mu_3$	−0·0000,3654	−0·0000,0430	−0·0000,012440	−0·0000,005184	−0·0000,002638
	μ_4	0·0000,0461	0·0000,0026	0·0000,000514	0·0000,000152	0·0000,000[illegible]
	${}_1\mu_4$	0·0000,0462	0·0000,0026	0·0000,000508	0·0000,000154	0·0000,000067
	β_1	1·61	1·81	1·88	1·91	1·93
	β_2	5·2	5·6	5·9	5·6	5·5
2	μ_1'	0·8642,3289	0·9362,9104	0·9583,8364	0·9880,[illegible]	0·9754,270546
	μ_2	0·0025,9457	0·0006,2664	0·0002,7472	0·0001,534270	0·0000,977034
	μ_3	−0·0000,8171	−0·0000,1137	−0·0000,0345	−0·0000,014671	−0·0000,007557
	${}_1\mu_3$	−0·0000,8178	−0·0000,1137	−0·0000,0345	−0·0000,014677	−0·0000,007557
	μ_4	0·0000,2312	0·0000,0146	0·0000,0029	0·0000,000801	0·0000,000381
	${}_1\mu_4$	0·0000,2312	0·0000,0146	0·0000,0029	0·0000,000812	0·0000,000383
	β_1	0·38	0·53	0·57	0·59	0·61
	β_2	3·4	3·7	3·9	3·8	4·0
3	μ_1'	0·7336,9741	0·8737,2072	0·9172,5148	0·9384,6566	0·9510,224795
	μ_2	0·0042,3560	0·0011,4861	0·0005,1075	0·0002,9459	0·0001,893284
	μ_3	−0·0000,8380	−0·0000,1767	−0·0000,0590	−0·0000,0262	−0·0000,013829
	${}_1\mu_3$	−0·0000,8401	−0·0000,1769	−0·0000,0590	−0·0000,0262	−0·0000,013826
	μ_4	0·0000,5399	0·0000,0427	0·0000,0089	0·0000,0029	0·0000,001226
	${}_1\mu_4$	0·0000,5403	0·0000,0428	0·0000,0089	0·0000,0029	0·0000,001214
	β_1	0·09	0·21	0·25	0·27	0·28
	β_2	3·0	3·2	3·3	3·4	3·4
4	μ_1'	0·5768,8665	0·7939,4237	0·8639,4783	0·8984,6442	0·9190,1507
	μ_2	0·0050,1034	0·0016,6539	0·0007,9361	0·0004,8861	0·0003,0008
	μ_2	−0·0000,2967	−0·0000,1968	−0·0000,0774	−0·0000,0369	−0·0000,0202
	${}_1\mu_3$	−0·0000,3099	−0·0000,1973	−0·0000,0775	−0·0000,0369	−0·0000,0202
	μ_4	0·0000,7256	0·0000,0850	0·0000,0198	0·0000,0067	0·0000,0029
	${}_1\mu_4$	0·0000,7262	0·0000,0850	0·0000,0198	0·0000,0067	0·0000,0029
	β_1	0·01	0·08	0·12	0·14	0·15
	β_2	2·9	3·1	3·1	3·2	3·2
5	μ_1'	0·4123,5896	0·7010,2826	0·8002,0843	0·8500,5217	0·8800,0945
	μ_2	0·0045,0478	0·0020,5600	0·0010,5521	0·0006,3313	0·0004,2033
	μ_3	+0·0000,3208	−0·0000,1587	−0·0000,0829	−0·0000,0437	−0·0000,0253
	${}_1\mu_3$	+0·0000,2882	−0·0000,1606	−0·0000,0831	−0·0000,0438	−0·0000,0253
	μ_4	0·0000,5903	0·0000,1262	0·0000,0341	0·0000,0123	0·0000,0055
	${}_1\mu_4$	0·0000,5897	0·0000,1262	0·0000,0339	0·0000,0124	0·0000,0055
	β_1	0·01	0·03	0·06	0·08	0·09
	β_2	2·9	3·0	3·1	3·1	3·1
6	μ_1'	0·2603,9120	0·5909,9719	0·7282,3688	0·7944,5984	0·8347,8956
	μ_2	0·0030,5059	0·0022,3214	0·0012,6522	0·0007,9285	0·0005,3942
	μ_3	+0·0000,4976	−0·0000,0779	−0·0000,0726	−0·0000,0446	−0·0000,0286
	${}_1\mu_3$	+0·0000,4559	−0·0000,0823	−0·0000,0734	−0·0000,0451	−0·0000,0281
	μ_4	0·0000,2830	0·0000,1472	0·0000,0480	0·0000,0185	0·0000,0095
	${}_1\mu_4$	0·0000,2807	0·0000,1472	0·0000,0481	0·0000,0191	0·0000,0089
	β_1	0·09	0·01	0·03	0·04	0·05
	β_2	3·0	3·0	3·0	2·9	3·3

TABLE II (*continued*)

q		$n=30$, $k=2$	$n=30$, $k=5$	$n=30$, $k=10$
1	μ_1'	0·9913,444323	0·9861,620246	0·9844,364077
	μ_2	0·0001,459732	0·0000,937517	0·0000,527977
	μ_3	−0·0000,047980	−0·0000,012438	−0·0000,003516
	${}_1\mu_3$	−0·0000,047992	−0·0000,012440	−0·0000,003516
	μ_4	0·0000,002947	0·0000,000514	0·0000,000136
	${}_1\mu_4$	0·0000,002947	0·0000,000508	0·0000,000124
	β_1	7·40	1·88	0·84
	β_2	13·8	5·9	4·9
2	μ_1'	0·9737,237167	0·9583,8364	0·9533,389957
	μ_2	0·0004,404662	0·0002,7472	0·0001,532142
	μ_3	−0·0000,141247	−0·0000,0345	−0·0000,009540
	${}_1\mu_3$	−0·0000,141253	−0·0000,0345	−0·0000,009535
	μ_4	0·0000,012312	0·0000,0029	0·0000,000813
	${}_1\mu_4$	0·0000,012309	0·0000,0029	0·0000,000787
	β_1	2·33	0·57	0·25
	β_2	6·4	3·9	3·5
3	μ_1'	0·9471,293171	0·9172,5148	0·9075,804094
	μ_2	0·0008,673760	0·0005,1975	0·0002,860673
	μ_3	−0·0000,264701	−0·0000,0590	−0·0000,015837
	${}_1\mu_3$	−0·0000,264137	−0·0000,0590	−0·0000,015851
	μ_4	0·0000,034381	0·0000,0090	0·0000,002570
	${}_1\mu_4$	0·0000,033682	0·0000,0090	0·0000,002566
	β_1	1·07	0·25	0·11
	β_2	4·6	3·3	3·1
4	μ_1'	0·9118,024417	0·8639,4783	0·8487,578437
	μ_2	0·0013,941771	0·0007,9361	0·0004,295258
	μ_3	−0·0000,391427	−0·0000,0774	−0·0000,019936
	${}_1\mu_3$	−0·0000,391348	−0·0000,0775	−0·0000,019982
	μ_4	0·0000,072591	0·0000,0198	0·0000,005613
	${}_1\mu_4$	0·0000,072561	0·0000,0198	0·0000,005636
	β_1	0·57	0·12	0·05
	β_2	3·7	3·1	3·0
5	μ_1'	0·8682,461741	0·8002,0843	0·7791,127138
	μ_2	0·0019,746075	0·0010,5521	0·0005,595144
	μ_3	−0·0000,493446	−0·0000,0829	−0·0000,020115
	${}_1\mu_3$	−0·0000,493538	−0·0000,0831	−0·0000,020241
	μ_4	0·0000,131257	0·0000,0341	0·0000,009414
	${}_1\mu_4$	0·0000,131262	0·0000,0339	0·0000,009434
	β_1	0·32	0·06	0·02
	β_2	3·4	3·1	3·0
6	μ_1'	0·8172,266960	0·7282,3688	0·7013,941584
	μ_2	0·0025,529007	0·0012,6522	0·0006,547817
	μ_3	−0·0000,543573	−0·0000,0726	−0·0000,016168
	${}_1\mu_3$	−0·0000,544455	−0·0000,0734	−0·0000,016440
	μ_4	0·0000,206107	0·0000,0480	0·0000,012833
	${}_1\mu_4$	0·0000,206344	0·0000,0481	0·0000,012845
	β_1	0·18	0·03	0·01
	β_2	3·2	3·0	3·0

TABLE II (*continued*)

	$q=1$, $k=2$, $n=10$	$q=1$, $k=2$, $n=30$	$q=1$, $k=2$, $n=50$
μ_1'	0·9719,147029	0·9913,444323	0·9848,854137
μ_2	0·0014,478736	0·0001,459732	0·0000,515182
μ_3	−0·0001,373323	−0·0000,047980	−0·0000,010221
${}_1\mu_3$	−0·0001,376674	−0·0000,047992	−0·0000,010222
μ_4	0·0000,242507	0·0000,002947	0·0000,000375
${}_1\mu_4$	0·0000,243786	0·0000,002947	0·0000,000376
β_1	6·21	7·40	7·64
β_2	11·6	13·8	14·1
	$q=2$, $k=2$, $n=10$	$q=2$, $k=2$, $n=30$	$q=2$, $k=2$, $n=50$
μ_1'	0·9122,379882	0·9737,237167	0·9845,495758
μ_2	0·0044,231601	0·0004,404682	0·0001,550848
μ_3	−0·0003,816694	−0·0000,141247	−0·0000,030326
${}_1\mu_3$	−0·0003,818601	−0·0000,141253	−0·0000,030334
μ_4	0·0000,997748	0·0000,012312	0·0000,001570
${}_1\mu_4$	0·0000,998432	0·0000,012309	0·0000,001584
β_1	1·68	2·33	2·47
β_2	5·1	6·4	6·5

It will be noted that there are sections common to the three parts of Table II. These have been included more than once in order to facilitate comparisons in any one part of the table.

for which $q = 1$ when discussing the adequacy of the Type I approximation to the distribution of l_1.

Consideration of an alternative form of approximation

For the larger values of n which have been considered it is seen that the distribution of l_1 is of small dispersion and is situated close to unity, the upper limit of possible values of l_1. In fact, in many of the cases dealt with, the mean is separated from the start of the curve by some twenty or thirty times the standard deviation. In these circumstances it seemed possible that a better approximation might be obtained by taking a Pearson Type I curve in the form

$$p(l_1) = \frac{\Gamma(m_1+m_2)}{(1-b)^{m_1+m_2-1}\,\Gamma(m_1)\,\Gamma(m_2)}(l_1-b)^{m_1-1}(1-l_1)^{m_2-1}, \text{ for } b \leqslant l_1 \leqslant 1. \quad \ldots(19)$$

There are now three assignable parameters; b, defining the start of curve, m_1 and m_2, which may be selected by equating the first three moments of the distribution (19) to those of the true distribution of l_1. As before, an idea of the degree of approximation may be obtained by comparing the fourth moment of l_1,

calculated from the mean, with that of the distribution (19). This procedure has been carried out in a few cases, the results obtained being set out in the upper part of Table III.

As far as can be judged by a comparison of moments, there is little difference between the approximations obtained by using a Type I curve in form (7), which is fitted by two moments, and the Type I curve in form (19), which is fitted by three moments. However, as we are primarily concerned with the probability levels of the distribution of l_1, the effect on these levels of the change in the method of approximation must be ascertained.

TABLE III

Comparison of distributions fitted by two and three moments

		$n=30$, $k=5$, $q=1$	$n=30$, $k=5$, $q=2$	$n=30$, $k=5$, $q=3$	$n=30$, $k=5$, $q=4$	$n=30$, $k=5$, $q=5$	$n=30$, $k=5$, $q=6$
$\mu_4 \times 10^8$	Of true distribution	5·14	29	90	198	341	480
	Of distribution (7)	5·08	29	90	198	339	481
	Of distribution (19)	5·08	29	90	198	339	480
5 % level	Of distribution (7)	0·9674	0·9280	0·8767	0·8149	0·7446	0·6678
	Of distribution (19)	0·9674	0·9280	0·8767	0·8149	0·7446	0·6677
1 % level	Of distribution (7)	0·9546	0·9110	0·8563	0·7915	0·7192	0·6411
	Of distribution (19)	0·9546	0·9110	0·8562	0·7915	0·7192	0·6413

The method of obtaining the 5 % and 1 % levels when the distribution is in form (7) has been given. When the second approximation is considered, the transformation

$$t = \frac{l_1 - b}{1 - b} \qquad \ldots\ldots(20)$$

gives the distribution of t as

$$p(t) = \frac{\Gamma(m_1+m_2)}{\Gamma(m_1)\,\Gamma(m_2)}\, t^{m_1-1}\,(1-t)^{m_2-1}, \text{ for } 0 \leqslant t \leqslant 1,$$

so that, as before, the tables of z may be utilized in calculating the levels. The lower part of Table III compares the levels given by the two methods of approximation.

It is seen that, for all practical purposes, the probability levels as given by the two forms of Type I curve may be taken as identical, so that nothing is to be gained by using the more complicated distribution (19). Furthermore, the agreement obtained strengthens the conviction that the Type I curve in form (7) really provides a good approximation to the true distribution. From now onward, therefore, we shall only be concerned with the approximate distribution given by (7).

5. Direct comparisons in cases where the true distribution of l_1 is known

When only two groups are considered the distribution of l_1 is known in the cases of one and two variates. These known distributions may be used to test the accuracy of the probability levels as obtained by using the Type I curve. If $q = 1$ we have seen that $l_1^2 = L_1$. P. P. N. Nayer (1936, p. 43) has shown that, when $k = 2$,

$$p(L_1) = \frac{\Gamma(n-1)}{2^{n-3}\left\{\Gamma\left(\frac{n-1}{2}\right)\right\}^2} L_1^{n-2}(1-L_1^2)^{-\frac{1}{2}}, \text{ for } 0 \leq L_1 \leq 1.$$

$$\text{Now,}\quad P\{l_1 < l_1^0\} = P\{L_1 < (l_1^0)^2\} = \frac{\Gamma(n-1)}{2^{n-3}\left\{\Gamma\left(\frac{n-1}{2}\right)\right\}^2}\int_0^{(l_1^0)^2} L_1^{n-2}(1-L_1^2)^{-\frac{1}{2}}\,dL_1.$$

Hence,
$$P\{l_1 < l_1^0\} = I_x(\tfrac{1}{2}(n-1), \tfrac{1}{2}), \text{ where } x = (l_1^0)^4. \qquad \text{......(21)}$$

When $q = 2$ and $k = 2$, Pearson & Wilks (1933) have proved that

$$P\{l_1 < l_1^0\} = \frac{\Gamma(2n-3)}{\Gamma(n-1)\,\Gamma(n-2)\,2^{2n-5}}\left\{(l_1^0)^{2n-4}\log_e\left(\frac{1+\sqrt{\{1-(l_1^0)^2\}}}{l_1^0}\right)\right\} + I_x(n-2, \tfrac{1}{2}),$$

where $x = (l_1^0)^2$. ......(22)

Thus if $k = 2$, relations (21) and (22) enable the probability integrals of l_1 ($q = 1$) and l_1 ($q = 2$) to be obtained by using *Tables of the Incomplete Beta Function*. In this manner the true probability associated with each of the 5 % and 1 % limits, obtained by using the Type I approximation, may be calculated. Results obtained are given in Table IV.

For the values of n considered it is seen, from Table IV, that the true probabilities are very close to the desired values 0·05 and 0·01. It should be pointed out that, as is shown in Table II, the six distributions considered above are unusually skew and leptokurtic compared with the remainder of the distributions, and moreover the agreement of true and approximate distribution, as judged by the moments, is not noticeably better for these curves than for the curves in general. Hence, for distributions which, judging from the values of β_1 and β_2, are certainly not in any way favoured, the limits set by using the approximation are found to give true probabilities near enough to the desired values for all practical purposes.

To summarize the salient points in the preceding sections, we have seen that:

(*a*) For a wide range of distributions close agreement between the third and fourth moments of the true and approximate distributions is obtained.

(*b*) Type I curves in form (19) fitted by three moments, which might be expected to give a better approximation, lead to probability levels which are for practical purposes the same as those given by the Type I curve fitted by two moments.

(c) In cases where the true distribution is known, the limits set by the approximation give values of the true probability near to those desired.

We may therefore say with some confidence that, having regard to the above considerations, the use of the Type I curve in form (7) as an approximation to the distribution of l_1 seems to be amply justified.

TABLE IV

	$q=1$, $k=2$, $n=10$	$q=1$, $k=2$, $n=30$	$q=1$, $k=2$, $n=50$
Type I 5% limit	0·8936	0·9669	0·9804
True probability	0·05005	0·05008	0·04950
Type I 1% limit	0·8235	0·9435	0·9664
True probability	0·00998	0·01001	0·01002
	$q=2$, $k=2$, $n=10$	$q=2$, $k=2$, $n=30$	$q=2$, $k=2$, $n=50$
Type I 5% limit	0·7814	0·9324	0·9600
True probability	0·05005	0·04997	0·04970
Type I 1% limit	0·6989	0·9034	0·9425
True probability	0·00999	0·00996	0·00998

6. The limiting form of the distribution of l_1

As was the case in the single variate problem (Neyman & Pearson, 1931), the distribution of l_1 in large samples may be obtained approximately from that of χ^2. $M_h(\lambda_1)$, the hth moment of the sampling distribution of $\lambda_1 = l_1^N$, may be obtained from (6) by replacing h by Nh. Using Stirling's approximation,

$$\Gamma(x) \sim \sqrt{(2\pi)}\, x^{x-\frac{1}{2}} e^{-x},$$

it may readily be shown that $M_h(\lambda_1)$ tends uniformly to $(1+h)^{-\frac{1}{2}f}$ as n tends to ∞, for all $h \geqslant 0$, where

$$f = \tfrac{1}{2}(k-1)\, q(q+1). \qquad \text{......(23)}$$

The distribution of χ^2 with f degrees of freedom is

$$p(\chi^2) = \{2^{\frac{1}{2}f}\, \Gamma(\tfrac{1}{2}f)\}^{-1} (\chi^2)^{\frac{1}{2}f-1} e^{-\frac{1}{2}\chi^2}, \text{ for } 0 \leqslant \chi^2 \leqslant \infty. \qquad \text{......(24)}$$

If we put $y = e^{-\frac{1}{2}\chi^2}$ (so that $0 \leqslant y \leqslant 1$), the hth moment of y is

$$\begin{aligned} M_h(y) &= \int_0^1 y^h p(y)\, dy = \{2^{\frac{1}{2}f}\, \Gamma(\tfrac{1}{2}f)\}^{-1} \int_0^\infty (\chi^2)^{\frac{1}{2}f-1} e^{-\frac{1}{2}\chi^2(1+h)}\, d(\chi^2) \\ &= (1+h)^{-\frac{1}{2}f}. \end{aligned}$$

Thus the corresponding moments of the distribution of y and the limiting form of the distribution of λ_1 are equal. The range of possible values of the variables being finite, i.e. from zero to unity, the equality of moments is sufficient to ensure that the distribution of y and the limiting distribution of λ_1 are identical.

Therefore, in large samples, $-2\log_e \lambda_1 = -2N \log_e l_1$ will be approximately distributed as χ^2 with f degrees of freedom. Thus if

$$l_1 = e^{-\chi^2/2N}, \qquad \ldots\ldots(25)$$

it follows that the probability levels for l_1 may be obtained by inserting in (25) the corresponding levels for χ^2. It had been hoped that this method of calculating the probability levels might prove satisfactory for moderately large values of n, so that the labour of calculating the first two moments of l_1 could have been avoided. In actual fact, however, this hope is not realized, as is shown by the results given in Table V.

TABLE V

q		$k=2$, $n=30$		$k=5$, $n=30$		$k=5$, $n=40$		$k=5$, $n=50$		$k=10$, $n=30$	
		5%	1%	5%	1%	5%	1%	5%	1%	5%	1%
1	(1) Level from Type I	0·9669	0·9435	0·9674	0·9546	0·9757	0·9662	0·9807	0·9731	0·9709	0·9629
	(2) Level from χ^2	0·9685	0·9462	0·9680	0·9567	0·9766	0·9673	0·9812	0·9738	0·9722	0·9645
	Difference (2)−(1) divided by $\sqrt{(\mu_2)}$	0·13	0·22	0·15	0·22	0·12	0·15	0·10	0·12	0·18	0·22
2	(1) Level from Type I	—	—	0·9280	0·9110	0·9463	0·9336	0·9573	0·9470		
	(2) Level from χ^2	—	—	0·9323	0·9163	0·9488	0·9366	0·9588	0·9489		
	Difference (2)−(1) divided by $\sqrt{(\mu_2)}$	—	—	0·26	0·32	0·20	0·24	0·15	0·19		
3	(1) Level from Type I	0·8913	0·8576	0·8767	0·8563	0·9079	0·8922	0·9265	0·9138	0·8783	0·8643
	(2) Level from χ^2	0·9004	0·8693	0·8857	0·8665	0·9130	0·8981	0·9298	0·9176	0·8867	0·8736
	Difference (2)−(1) divided by $\sqrt{(\mu_2)}$	0·31	0·40	0·39	0·45	0·30	0·34	0·24	0·28	0·30	0·35
4	(1) Level from Type I	—	—	0·8140	0·7915	0·8610	0·8428	0·8887	0·8737		
	(2) Level from χ^2	—	—	0·8304	0·8087	0·8699	0·8528	0·8945	0·8804		
	Difference (2)−(1) divided by $\sqrt{(\mu_2)}$	—	—	0·55	0·61	0·41	0·47	0·33	0·38		
5	(1) Level from Type I	0·7884	0·7476	0·7446	0·7192	0·8067	0·7867	0·8446	0·8281		
	(2) Level from χ^2	0·8119	0·7750	0·7684	0·7448	0·8206	0·8018	0·8537	0·8380		
	Difference (2)−(1) divided by $\sqrt{(\mu_2)}$	0·53	0·62	0·73	0·79	0·55	0·60	0·44	0·48		
6	(1) Level from Type I	—	—	0·6678	0·6411	0·7462	0·7244	0·7935	0·7761	—	—
	(2) Level from χ^2	—	—	0·7014	0·6769	0·7665	0·7463	0·8083	0·7313	—	—
	Difference (2)−(1) divided by $\sqrt{(\mu_2)}$	—	—	0·94	1·01	0·72	0·78	0·64	0·65	—	—

It is evident from inspection of Table V that, as would be expected, the agreement between the levels, as given by the Type I and χ^2 approximations, improves with increase of n, whereas it slowly becomes worse as k is increased. However, as q increases there is such a rapid deterioration in the accuracy of the approximation that this method cannot be justifiably employed, except for large values of n.

It may however be usefully noted that the levels obtained from the χ^2 transformation are always above those set by the Type I curve, so that if the hypothesis is not rejected when the level given by χ^2 is employed it certainly would not be rejected if the more accurate Type I level had been calculated.

7. Empirical relations for m_1 and m_2

When using the test it is found that the calculation of the first and second moments of l_1 accounts for the major part of the labour of computation which is involved. It therefore seemed desirable to attempt to obtain empirical relations giving m_1 and m_2 in terms of n, k and q, such that the use of these values will lead to probability levels of l_1 sufficiently accurate for practical purposes.

We have seen that $M_h(\lambda_1)$ tends uniformly to $(1+h)^{-\frac{1}{2}f}$ as n tends to ∞. Put $l_1' = \lambda_1^{1/M}$, where M is any positive number. Then $M_h(l_1')$, the hth moment of l_1' tends uniformly to $(1+h/M)^{-\frac{1}{2}f}$ as $N \to \infty$. If a Type I curve be fitted to the distribution of l_1', in a way similar to the one employed when approximating to the distribution of l_1, the exponent of the power of $(1-l_1')$ will be given by

$$m_2' = \frac{\{1-M_1(l_1')\}\{M_1(l_1')-M_2(l_1')\}}{M_2(l_1')-\{M_1(l_1')\}^2},$$

and this may be considered as a function $\psi(M, N)$, of M and N.

Now,

$$\lim_{N\to\infty} \psi(M, N) = \lim_{N\to\infty} m_2' = \frac{\{1-(1+1/M)^{-\frac{1}{2}f}\}\{(1+1/M)^{-\frac{1}{2}f}-(1+2/M)^{-\frac{1}{2}f}\}}{(1+2/M)^{-\frac{1}{2}f}-(1+1/M)^{f}} = g(M),$$

the limit being approached uniformly for all positive M, and $\lim_{M\to\infty} g(M) = \frac{1}{2}f$.

Hence
$$\lim_{N\to\infty} \psi(N, N) = \lim_{N\to\infty} m_2 = \tfrac{1}{2}f = 0{\cdot}25(k-1)\,q(q+1).$$

Values of m_1 and m_2 which have been calculated in a number of cases are given in Table VI. The values of m_2 indicate that the limit $\frac{1}{2}f$ is approached rapidly with increase of n, and it seems probable that, for $n \geqslant 20$, no error of practical importance will be made if we substitute $\frac{1}{2}f$ for m_2.

Methods of approximating to m_1 must now be considered. U. S. Nair (1938, p. 285) has shown that the distribution of l_1 may be reduced to the form

$$p(l_1) = p_1(l_1)\,p_2(l_1),$$

where
$$p_1(l_1) = \text{const}\; l_1^{\frac{1}{2}k(n-q)-1}$$

and
$$p_2(l_1) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} l_1^{-t}\, k^{\frac{1}{2}qt} \prod_{j=1}^{q}\left\{\frac{(\Gamma(\frac{1}{2}(q-j+t/k)))^k}{\Gamma(\frac{1}{2}(k(q-1)+1-j+t))}\right\}dt,$$

TABLE VI

Values of m_1 and m_2

	$q=1$	$q=2$	$q=3$	$q=4$	$q=5$	$q=6$
$k=2$ $n=10$	17·351 0·501	15·599 1·501				
$k=2$ $n=30$	57·282 0·500	55·588 1·500	53·732 2·999	51·683 4·999	49·432 7·501	46·997 10·511
$k=2$ $n=50$	97·269 0·500	95·586 1·500				
$k=5$ $n=10$	42·097 2·005	38·219 6·004	33·111 12·018	27·527 20·189	21·769 31·022	16·178 45·953
$k=5$ $n=20$	92·501 2·001	88·188 6·000	83·053 12·004	77·190 20·030	70·758 30·177	63·912 42·[illegible]
$k=5$ $n=30$	142·559 2·000	138·177 6·000	133·033 12·001	127·094 20·014	120·439 30·071	113·187 42·239
$k=5$ $n=40$	192·543 2·000	188·178 6·000	183·014 12·000	177·043 20·008	170·285 30·038	162·830 42·127
$k=5$ $n=50$	242·532 2·000	238·175 6·000	233·020 12·001	227·017 20·005	220·191 30·023	212·689 42·075
$k=10$ $n=30$	284·689 4·501	275·836 13·501	265·205 27·006	252·811 45·049	238·861 67·719	223·649 95·214
$k=25$ $n=30$	711·207 12·004	688·901 36·007	...	...		

In each cell m_1 is the upper and m_2 the lower number.

so that $p_2(l_1)$ is independent of n. Now if $\frac{1}{2}f$ replaces m_2, the Type I distribution (7) takes the form

$$p(l_1) = \text{const}\; l_1^{m_1-1}(1-l_1)^{\frac{1}{2}f-1},$$

so that it is expressed as the product of two factors $l_1^{m_1-1}$ and $(1-l_1)^{\frac{1}{2}f-1}$, the second of which is independent of n. In these circumstances it seemed possible that m_1 might perhaps be replaced by $k(n-q)$. Therefore putting

$$m_1 = k(n-q) \qquad \ldots\ldots(26)$$

and

$$m_2 = \tfrac{1}{2}f = 0{\cdot}25(k-1)q(q+1), \qquad \ldots\ldots(16)\ bis$$

the 5 % and 1 % levels of l_1 were calculated in a number of cases and compared with those given by using the values of m_1 and m_2 obtained from the moments. The results given in Table VII indicate that, although the agreement in cases where $q = 1$ is satisfactory, for larger values of q the discrepancies become too large to enable this method of approximation to be used in practice.

TABLE VII

Comparison of levels given by Method 1 and Method 2

		$q=1$		$q=3$		$q=5$	
	Method	5%	1%	5%	1%	5%	1%
$k=2$ $n=30$	1 2	0·967 0·967	0·944 0·944	0·891 0·892	0·858 0·858	0·788 0·791	0·748 0·750
$k=5$ $n=10$	1 2	0·896 0·901	0·857 0·864	0·621 0·635	0·570 0·585	0·304 0·346	0·263 0·304
$k=5$ $n=20$	1 2	0·950 0·952	0·931 0·933	0·814 0·818	0·784 0·788	0·624 0·640	0·590 0·607
$k=5$ $n=30$	1 2	0·967 0·968	0·955 0·955	0·877 0·878	0·856 0·858	0·745 0·752	0·719 0·728
$k=5$ $n=40$	1 2	0·976 0·976	0·966 0·967	0·908 0·909	0·892 0·893	0·807 0·811	0·787 0·792
$k=5$ $n=50$	1 2	0·981 0·981	0·973 0·973	0·927 0·927	0·914 0·915	0·845 0·848	0·828 0·831
$k=10$ $n=30$	1 2	0·971 0·970	0·963 0·962	0·878 0·880	0·864 0·866	— —	— —

Method 1. m_1 and m_2 obtained from the first two moments of l_1.
Method 2. m_1 and m_2 given by the empirical relations

$$m_1 = k(n-q), \quad m_2 = 0{\cdot}25(k-1)\,q(q+1).$$

It is clear that the deviations are due almost entirely to an inadequate approximation to m_1, although the expression does not appear to differ from m_1 in such a marked manner that its use as a basis of approximation must be rejected entirely. Thus it was thought that some slight correcting term might be subtracted from $k(n-q)$ to provide a better approximation to m_1. With this end in view values of the necessary correction $C = k(n-q) - m_1$ were calculated and are given in Table VIII.

TABLE VIII

Values of correction $C = k(n-q) - m_1$

	$q=1$	$q=2$	$q=3$	$q=4$	$q=5$	$q=6$
$k=2,\ n=10$ $k=2,\ n=30$ $k=2,\ n=50$	0·65 0·72 0·73	0·40 0·41 0·41	0·27	0·32	0·57	1·08
$k=5,\ n=10$ $k=5,\ n=20$ $k=5,\ n=30$ $k=5,\ n=40$ $k=5,\ n=50$	2·30 2·41 2·44 2·46 2·47	1·72 1·81 1·82 1·82 1·82	1·89 1·95 1·97 1·99 1·98	2·47 2·80 2·91 2·96 2·98	3·23 4·24 4·56 4·71 4·81	3·82 6·09 6·81 7·17 7·40
$k=10,\ n=30$ $k=25,\ n=30$	5·31 13·79	4·16 11·10	4·79	7·19	11·11	16·35

When $n \geqslant 20$, the variation of C with n, for fixed k and q, is small, so that in seeking a simple empirical relation for C, functions of k and q only were considered. If, for constant q, C be plotted against k, it is found that the relation between them is nearly linear and, moreover, C may be taken as proportional to $(k-1)$. Assuming therefore that $C = (k-1)\theta_q$, values of θ_q for different q may be calculated. On plotting θ_q against q, the form of the resulting curve suggests that, over the range of values of q which are considered, θ_q will be given with sufficient accuracy by a quadratic form in q. Making this assumption the coefficients of the quadratic were adjusted by trial and error until the expression which resulted seemed likely to give satisfactory results in most cases. The correcting term was thus tentatively obtained in the form

$$C = 0{\cdot}01(k-1)(90-39q+9q^2).$$

The final replacement for m_1, as previously given, is therefore

$$m_1 = k(n-q) - 0{\cdot}01(k-1)(90-39q+9q^2), \qquad \ldots\ldots(15)\ bis$$

and as before $\quad m_2 = 0{\cdot}25(k-1)\,q(q+1). \qquad \ldots\ldots(16)\ bis$

It is realized that the reader may have some doubts as to the validity of the empirical methods by which this approximation for m_1 was obtained. In order therefore to establish this approximation as a satisfactory practical method, it is necessary to calculate for a wide range of cases the 5 % and 1 % levels given by the approximation. These may then be compared with the levels obtained when values of m_1 and m_2 calculated from the moments are used. This has been done in a number of cases, the results being given in Table IX.

Inspection of this table shows that when $n \geqslant 20$ the levels obtained by using values of m_1 and m_2 as given by (15) and (16) agree with those resulting

TABLE IX

Comparison of levels given by Method 1 and Method 3

			$q=1$		$q=2$		$q=3$		$q=4$		$q=5$		$q=6$	
k	n	Method	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%
2	10	1	0·894	0·824	0·781	0·699	—	—	—	—	—	—	—	—
		3	0·894	0·824	0·780	0·698	—	—	—	—	—	—	—	—
2	30	1	0·967	0·944	0·932	0·903	0·891	0·858	0·843	0·805	0·788	0·748	0·728	0·685
		3	0·967	0·944	0·932	0·903	0·891	0·857	0·842	0·804	0·786	0·745	0·725	0·681
2	50	1	0·980	0·966	0·960	0·943	—	—	—	—	—	—	—	—
		3	0·980	0·966	0·960	0·943	—	—	—	—	—	—	—	—
5	10	1	0·896	0·857	0·772	0·724	0·621	0·570	0·458	0·409	0·304*	0·263*	0·174*	0·145*
		3	0·896	0·857	0·771	0·723	0·619	0·567	0·454	0·405	0·202	0·251	0·146	0·118
5	20	1	0·950	0·931	0·891	0·865	0·814	0·784	0·724	0·691	0·624	0·590	0·521	0·488
		3	0·950	0·931	0·890	0·865	0·814	0·784	0·723	0·690	0·624	0·590	0·520	0·486
5	30	1	0·967	0·955	0·928	0·911	0·877	0·856	0·815	0·792	0·745	0·719	0·668	0·641
		3	0·967	0·955	0·928	0·911	0·877	0·856	0·815	0·791	0·745	0·719	0·668	0·641
10	30	1	0·971	0·963	0·931	0·920	0·878	0·864	0·813	0·797	—	—	—	—
		3	0·971	0·963	0·931	0·920	0·878	0·864	0·813	0·797	—	—	—	—
25	30	1	0·975	0·970	0·936	0·930	—	—	—	—	—	—	—	—
		3	0·975	0·970	0·936	0·929	—	—	—	—	—	—	—	—

Method 1. m_1 and m_2 obtained from the first two moments of l_1.
Method 3. m_1 and m_2 given by the empirical relations

$$m_1 = k(n-q) - 0{\cdot}01(k-1)(90 - 39q + 9q^2), \quad m_2 = 0{\cdot}25(k-1)\,q(q+1).$$

from the use of values of m_1 and m_2 calculated from the moments with sufficient accuracy for practical purposes. Even when n is as small as 10 the only serious discrepancies, marked with an asterisk, occur when there are as many as five or six variables.

The range of values of m_1 and m_2 of the distributions considered in the last table may be found by consulting Table VI. From the practical viewpoint, it must be noted that when both m_1 and m_2 exceed 60 it becomes impossible to interpolate in the z-tables without running the risk of making appreciable errors. In these cases an approximation, due to Fisher (1938*b*, § 41), may be used. When z has degrees of freedom f_1 and f_2 which are both large and $f = 2f_1f_2/(f_1+f_2)$ is their harmonic mean, the 5 % level of z is given approximately by

$$z_{\cdot 05} = \frac{1{\cdot}6449}{\sqrt{(f-1)}} - 0{\cdot}7843\left(\frac{1}{f_1} - \frac{1}{f_2}\right)$$

and the 1 % level by

$$z_{\cdot 01} = \frac{2{\cdot}3263}{\sqrt{(f-1)}} - 1{\cdot}235\left(\frac{1}{f_1} - \frac{1}{f_2}\right).$$

8. Some consideration of the case of unequal samples

We have so far considered k equal samples each of size n, but in the general case there may be k samples of sizes n_t $(t = 1, 2, \ldots, k)$. The appropriate criterion in this case is

$$l_1 = \lambda_1^{1/N} = \sqrt[q]{\left(\prod_{t=1}^{k}\left\{\frac{|v_{sut}|}{|v_{su}|}\right\}^{n_t/N}\right)},$$

where now $\quad v_{sut} = \dfrac{1}{n_t}\sum_{i=1}^{n_t}(x_{sti}-x_{st.})(x_{uti}-x_{ut.}), \quad v_{su} = \dfrac{1}{N}\sum_{t=1}^{k} n_t v_{sut}$

and $\quad x_{st.} = \dfrac{1}{n_t}\sum_{i=1}^{n_t} x_{sti}, \quad N = \sum_{t=1}^{k} n_t.$

It is seen that l_1^2 is the ratio of a weighted geometric mean of the generalized variances to a determinant of order q in which the element in the sth row and uth column is v_{su}, the weighted arithmetic mean of $v_{su1}, v_{su2}, \ldots, v_{suk}$, the weight given in each case being the corresponding sample size. When considering single-variate criteria, Bartlett (1937) suggested that a better test will be obtained if each variance is weighted with the number of degrees of freedom it possesses rather than with the sample size, and more recent unpublished work confirms this contention. Thus there is a possibility that in the general case, where more than one variate is considered, some adjustment of the weighting will yield a criterion which will more frequently detect the falsehood of the hypothesis H_0 when in fact some alternative hypothesis is true. That is to say, the modified test might prove to be more powerful, in the sense of Neyman & Pearson (1936), with regard to a certain set of alternative hypotheses. However, it is likely that such a modification will only have an appreciable effect when some, at least, of the samples are small.

The preceding reasoning concerning the limiting form of the distribution of l_1 may be shown to apply whether the samples are of the same size or not. The distribution of l_1, as approximated to by means of the χ^2 transformation, depends only on N, k and q, and not on the size of individual samples. It therefore seems possible that with large samples, which although unequal do not vary greatly in size, a reasonable approximation to the true levels may be obtained by using the χ^2 transformation of equation (25).

The accuracy of the χ^2 method of approximation may be improved by following a procedure similar to that used by Bartlett (1937) when dealing with a single variate. In the general case, the hth moment of λ_1 is given by

$$M_h(\lambda_1) = \prod_{t=1}^{k}\left\{\left(\frac{N}{n_t}\right)^{\frac{1}{2}qhn_t}\prod_{i=1}^{q}\frac{\Gamma\left(\frac{n_t(1+h)-i}{2}\right)}{\Gamma\left(\frac{n_t-i}{2}\right)}\right\}\cdot\prod_{i=1}^{q}\left\{\frac{\Gamma\left(\frac{N-k+1-i}{2}\right)}{\Gamma\left(\frac{N(1+h)-k+1-i}{2}\right)}\right\}.$$

In the limit $-2\log_e\lambda_1$ is distributed as χ^2 with f degrees of freedom and our purpose is to find a correcting factor G, a function of n_t, k and q which only involves terms of order $(n_t)^{-1}$, such that $-2G^{-1}\log_e\lambda_1$ has a mean value differing

from f, the mean value of χ^2, by terms of order $(n_t)^{-2}$. Thus Gf must be identical with the sum of terms of order $(n_t)^{-1}$ in the coefficient of h in the expansion, in powers of h, of $\log_e\{M_{-2h}(\lambda_1)\}$. Using Stirling's approximation we obtain

$$G = 1+\frac{1}{f}\left\{\sum_{t=1}^{k}\sum_{i=1}^{q}\left(\frac{i^2}{2n_t}+\frac{i}{n_t-i}+\frac{n_t}{3(n_t-i)^2}\right) - \sum_{i=1}^{q}\left(\frac{(k+i-1)^2}{2N}+\frac{k+i-1}{N-k+1-i}+\frac{N}{3(N-k+1-i)^2}\right)\right\}.$$

If now $-2G^{-1}\log_e\lambda_1$ be referred to the χ^2 distribution with f degrees of freedom, approximate levels for l_1 may be obtained from the corresponding levels for χ^2 by using the relation

$$l_1 = \exp\left(-\frac{G\chi^2}{2N}\right). \qquad \text{......(27)}$$

The degree of accuracy provided by this method is indicated in Table X, which compares, for some typical cases, the levels given by (27) with those obtained by fitting a Type I curve.

TABLE X

			$q=1$	$q=2$	$q=3$	$q=4$	$q=5$	$q=6$
$k=5$, $n_t=10$, $(t=1, \ldots, 5)$	5% level	From Type I	0·896	0·772	0·621	0·458	0·304	0·174
		From relation (27)	0·896	0·774	0·628	0·473	0·328	0·206
	1% level	From Type I	0·857	0·724	0·570	0·409	0·263	0·145
		From relation (27)	0·857	0·726	0·577	0·425	0·288	0·176
$k=5$, $n_t=20$, $(t=1, \ldots, 5)$	5% level	From Type I	0·950	0·891	0·814	0·724	0·624	0·521
		From relation (27)	0·950	0·890	0·813	0·724	0·628	0·528
	1% level	From Type I	0·931	0·865	0·784	0·691	0·590	0·488
		From relation (27)	0·931	0·866	0·783	0·692	0·594	0·495

It is apparent from Tables V and X that the crude χ^2 approximation has been considerably improved by the use of the correcting factor G. Although examples in which the k samples are of *unequal* sizes have not been given in Table X, it is probable that, if the smallest sample contains at least twenty individuals, the application of the above method will give a fairly satisfactory result.

Mention may be made in passing of a similar approximation which is found to give close agreement with the distribution of the comprehensive likelihood-ratio criterion for testing one of the hypotheses concerning the values of the means of populations referred to in the introduction to this paper. It is assumed that the hypothesis H_0 regarding the equality of variances and covariances is true, and it is desired to test whether the corresponding means of the q characters are

the same in all k populations. In the case of two samples ($k = 2$), the test is equivalent to use of Hotelling's generalized T. If

$$a_{su} = \frac{1}{N}\sum_{t=1}^{k}\sum_{i=1}^{n_t}(x_{sti}-x_{s..})(x_{uti}-x_{u..}),$$

where

$$x_{s..} = \frac{1}{N}\sum_{t=1}^{k}\sum_{i=1}^{n_t} x_{sti},$$

the appropriate comprehensive likelihood criterion was given by Wilks in the form

$$W = \frac{|v_{su}|}{|a_{su}|}.$$

The distribution of W may be expressed simply in the following three cases:

(a) When $q = 1$, $p(W) = \left\{B\left(\frac{N-k}{2}, \frac{k-1}{2}\right)\right\}^{-1} W^{\frac{1}{2}(N-k-2)}(1-W)^{\frac{1}{2}(k-3)}$.*

(b) When $q = 2$, $p(\sqrt{W}) = \{B(N-k-1, k-1)\}^{-1}(\sqrt{W})^{N-k-2}(1-\sqrt{W})^{k-2}$.

(c) When $k = 2$, $p(W) = \left\{B\left(\frac{N-q-1}{2}, \frac{q}{2}\right)\right\}^{-1} W^{\frac{1}{2}(N-q-3)}(1-W)^{\frac{1}{2}(q-2)}$.†

In other cases, following the previous procedure, it may be shown that approximate levels for W may be obtained by using the relation

$$W = \exp\left(-\frac{G'\chi^2}{N}\right),$$

where

$$G' = 1 + \frac{1}{q(k-1)}\sum_{i=1}^{q}\left\{\frac{i(k-2)}{N-i} + \frac{N(3(k-1)^2-2)}{6(N-i)^2} + \frac{k+i-1}{N-k+1-i} + \frac{N}{3(N-k+1-i)^2}\right\},$$

and χ^2 has $q(k-1)$ degrees of freedom. Calculations which have been made for certain examples show that this method of approximation will give the levels of W with sufficient accuracy for most practical purposes if the ratio N/k is not less than 10.

Summary

The main object of this paper has been to put into easier form for application the likelihood ratio criterion l_1 for testing the hypothesis that all the corresponding variances and covariances in a number, k, of multivariate q-variable normal populations are identical. In the case where the samples from each population are of the same size ($n_t = n$, $t = 1, \ldots, k$) empirical formulae have been given, using which 5 % and 1 % probability levels for l_1 may be obtained readily from the levels of R. A. Fisher's z; alternatively the *Tables of the Incomplete Beta-*

* When $q = 1$, W is just the ratio of two sums of squares, familiar in the analysis of variance.

† $W = \left(1+\frac{T^2}{N-2}\right)^{-1}$.

Function may be used. The conditions under which these formulae may be safely employed are discussed.

If the samples are large and $N = \sum_{t=1}^{k} n_t$, $-2N \log_e l_1$ is distributed approximately as χ^2 with degrees of freedom $f = \frac{1}{2}(k-1)\, q(q+1)$. The test is then independent of the individual values of n_t, which may now differ. If there are many variables the sample sizes must, however, be very large for this transformation to be justifiable.

A more accurate approximation, of the type suggested by M. S. Bartlett, in the case of a single variable ($q = 1$), has also been considered. The corrective term is, however, rather troublesome to calculate unless again $n_t = n$ ($t = 1, \ldots, k$).

In conclusion I wish to thank Prof. E. S. Pearson for much assistance in the preparation of this paper.

Additional note

U. S. Nair has recently obtained, on a theoretical basis, an approximation to the distribution of l_1, in the case of k equal samples, somewhat similar to the one given as Method 3 on pp. 50–51.

He finds that the distribution of l_1 may be taken approximately as

$$p(l_1) = \text{const}\ l_1^{k(n-q)-C''-1}\,(1-l_1)^{\frac{1}{2}f-1},$$

where
$$C'' = \frac{1}{24}(k-1)\,(3q^2-13q+28) - \frac{1}{6}(2q-5) - \frac{k+1}{3(q+1)}.$$

As far as we have been able to check, this approximation and Method 3 seem to be adequate for the same values of n and appear to give levels which are substantially the same.

REFERENCES

BARTLETT, M. S. (1937). *Proc. Roy. Soc.* A, **160**, 268–82.
BISHOP, D. J. & NAIR, U. S. (1939). *J.R. Statist. Soc., Suppl.* **6**, 89–99.
FISHER, R. A. (1938*a*). *Ann. Eugen., Lond.*, **8**, 376–86.
—— (1938*b*). *Statistical Methods for Research Workers*, 7th edition. London: Oliver and Boyd.
HSU, P. L. (1938). *Ann. Math. Statist.* **9**, 231–43.
LAWLEY, D. N. (1938*a*). *Biometrika*, **30**, 180–87.
—— (1938*b*). *Biometrika*, **30**, 467–9.
NAIR, U. S. (1938). *Biometrika*, **30**, 274–94.
NAYER, P. P. N. (1936). *Statist. Res. Mem.* **1**, 38–51.
NEYMAN, J. & PEARSON, E. S. (1931). *Bulletin de l'Académie Polonaise des Sciences et des Lettres*, A, 460–81.
—— —— (1936). *Statist. Res. Mem.* **1**, 1–37.
PEARSON, E. S. (1922). *Tracts for Computers*, **8**. Cambridge University Press.
PEARSON, E. S. & WILKS, S. S. (1933). *Biometrika*, **25**, 353–78.
PEARSON, K. (1934). *Tables of the Incomplete Beta-Function*. Biometrika Office, University College, London.
PITTARD, E. (1909). *Crania Helvetica I. Les Crânes Valaisans de la Vallée du Rhône*. Extrait du tome 20 des Mémoires de l'Institut national genevois.
WILKS, S. S. (1932). *Biometrika*, **24**, 471–94.

OBSERVED AND THEORETICAL RATIOS IN MENDELIAN INHERITANCE

By E. ROBERTS, W. M. DAWSON* and MARGARET MADDEN*
University of Illinois

According to the present conception of Mendelian inheritance genes are located on chromosomes, and if not located on the same chromosome are independent in inheritance. During the formation of reproductive cells the number of chromosomes is reduced to one-half the number common to the species or one-half the number found in the somatic cells. As a result of fertilization the original number is restored.

An allelomorphic pair of genes are genes occupying the same relative loci on homologous chromosomes, one of which came from the father and the other from the mother. The members (alleles) of an allelomorphic pair under usual conditions separate, going into different reproductive cells or gametes. When located on different chromosomes the result is as many different kinds of germ cells as the number of possible combinations of the genes, except that members of an allelomorphic pair are not found in the same gamete. For any number of allelomorphic pairs the number of different kinds of gametes is 2^n, in which n is the number of allelomorphic pairs. In case of dominance and recessiveness the number of visible classes or phenotypes formed among the offspring as the result of random fertilization is also 2^n.

The purpose of this paper is to put on record the ratios obtained in several mammalian crosses made in genetic studies in the Laboratory of Animal Genetics, College of Agriculture, University of Illinois, during a period of several years.

Daily inspection for birth of young was made and number of young recorded. Usually two classifications were made, one at 12–18 days of age and the second at the time of weaning, which in the case of the mice and rats was at about 28 days. Rabbits were weaned at an age of about 8 weeks. Colours and colour patterns can be easily classified as soon as the pelage is well formed. Dark eye and pink eye are distinguishable at birth and were recorded when the birth records were made. In hypotrichotic rats the hair is normal in appearance until between 16 and 20 days of age when depilation begins, first noticeable around the head.

In our records litters were classified as complete and incomplete. A complete litter is one in which all the young were classified and an incomplete litter one in which one or more had died before classification. Ratios among incomplete litters did not differ significantly from ratios among complete litters, and for this reason they are combined in the tables.

* Formerly Assistant in Animal Husbandry.

The symbols used and the characters which they designate are as follows:

Rats

A = agouti, **a** = non-agouti.

H = self colour (entirely coloured), **h** = hooded.

R = dark eye, **r** = red eye.

C = colour, **c** = albinism.

$\mathbf{H_y}$ = haired, $\mathbf{h_y}$ = hypotrichosis (hairless).

D = intense colour, **d** = dilute colour.

Mice

A = agouti, **a** = non-agouti.

$\mathbf{P_1}$ = dark eye, $\mathbf{p_1}$ = pink eye.

$\mathbf{P_2}$ = dark eye, $\mathbf{p_2}$ = pink eye (a second gene causing pink eye).

B = black, **b** = brown.

Y = yellow, **y** = non-yellow (black or brown).

Rabbits

A = agouti, **a** = non-agouti.

E = extension, **e** = non-extension (yellow).

For obtaining probabilities of ratios with two classes the formula

$$E = \pm 0{\cdot}6745\sqrt{(npq)},$$

in which n is the number of individuals, p the observed proportion of one class and q the proportion of the other class, was used. For ratios having three or more terms, the fit of the observed to the theoretical ratio was determined by Pearson's formula

$$\chi^2 = S\left\{\frac{(m_1' - m_r)^2}{m_r}\right\};$$

m_1' is the observed number in a class, m_r the theoretical or expected number, and S is the sum. P is obtained from Pearson's *Tables for Statisticians and Biometricians*.

1. Matings of the type Aa × aa. Theoretical ratio of 1 : 1.

$$\left.\begin{array}{l} Aa \rightarrow A + a \\ aa \rightarrow a + a \end{array}\right\} \text{gametes.}$$

$$Aa + aa \text{ zygotes.}$$

TABLE I

Mating		Dominant	Recessive	Deviation	Probable error	*D/E*
Rats						
Aa × aa	O E	690 661·5	633 661·5	28·5	12·26	2·32
Hh × hh	O E	552 500	448 500	52·0	10·61	4·90
Rr × rr	O E	418 409·5	401 409·5	8·5	9·65	0·88
Cc × cc	O E	168 172·5	177 172·5	4·5	6·26	0·72
$H_yh_y \times h_yh_y$	O E	745 732·5	720 732·5	12·5	12·90	0·97
Dd × dd	O E	669 622·5	576 622·5	46·5	11·86	3·92
Mice						
Aa × aa	O E	1906 1867	1828 1867	39·0	20·61	1·89
$P_1p_1 \times p_1p_1$	O E	1898 1867	1836 1867	31·0	20·61	1·50
Bb × bb	O E	1842 1867	1892 1867	25·0	20·61	1·21
$P_2p_2 \times p_2p_2$	O E	374 345	316 345	29·0	8·83	3·28
Yy × yy	O E	259 264·5	270 264·5	5·5	7·70	0·71
Rabbits						
Ee × ee	O E	27 25·5	24 25·5	1·5	2·40	0·62
Aa × aa	O E	40 36	32 36	4·0	2·84	1·41

2. Matings of the type **Aa** × **Aa**, producing two phenotypes in the ratio of 3 dominants to 1 recessive.

$$\left.\begin{array}{l}\mathrm{Aa} \rightarrow \mathrm{A} + \mathrm{a} \\ \mathrm{Aa} \rightarrow \mathrm{A} + \mathrm{a}\end{array}\right\} \text{gametes.}$$

$$\overline{1\mathrm{AA} + 2\mathrm{Aa} + 1\mathrm{aa}} \text{ zygotes.}$$

Since A is dominant a ratio of 3 : 1 is expected.

TABLE II

Mating		Dominant	Recessive	Deviation	Probable error	*D/E*
Rats						
Aa × Aa	O	607	185	13·00	8·03	1·62
	E	594	198			
Hh × Hh	O	622	227	14·75	8·70	1·70
	E	636·75	212·25			
Rr × Rr	O	389	128	1·25	6·62	0·19
	E	387·75	129·25			
Cc × Cc	O	946	324	6·50	10·47	0·62
	E	952·5	317·5			
Hyhy × Hyhy	O	1305	435	0·00	12·18	0·00
	E	1305	435			
Dd × Dd	O	605	204	1·75	8·33	0·21
	E	606·75	202·25			
Mice						
$P_2p_2 \times P_2p_2$	O	471	141	12·00	7·03	1·71
	E	459	153			
Rabbits						
Ee × Ee	O	125	33	6·50	3·46	1·88
	E	118·5	39·5			

3. Matings of the type **AaBb × aabb** resulting in a theoretical ratio of 1 : 1 : 1 : 1.

TABLE III

Matings		Phenotypes				χ^2	P
Rats							
		AHy	Ahy	aHy	ahy		
AaHyhy × aahyhy	O	187	175	165	174	1·396	0·711
	E	175·25					
		HHy	Hhy	hHy	hhy		
HhHyhy × hhhyhy	O	194	187	146	158	9·219	0·027
	E	171·25					
		RHy	Rhy	rHy	rhy		
RrHyhy × rrhyhy	O	146	150	145	147	0·095	0·992
	E	147					
		DHy	Dhy	dHy	dhy		
DdHyhy × ddhyhy	O	158	166	144	138	3·241	0·360
	E	151·5					
		CHy	Chy	cHy	chy		
CcHyhy × cchyhy	O	13	13	16	12	0·668	0·880
	E	13·5					
		AD	Ad	aD	ad		
AaDd × aadd	O	181	164	187	159	2·967	0·398
	E	172·75					
		HD	Hd	hD	hd		
HhDd × hhdd	O	187	149	133	133	12·937	0·005
	E	150·5					
		RD	Rd	rD	rd		
RrDd × rrdd	O	184	160	175	161	2·364	0·507
	E	170					
		AR	Ar	aR	ar		
AaRr × aarr	O	150	146	147	143	0·172	0·982
	E	146·5					
		HR	Hr	hR	hr		
HhRr × hhrr	O	167	159	134	131	6·543	0·089
	E	147·75					
		AH	Ah	aH	ah		
AaHh × aahh	O	175	136	164	147	5·720	0·128
	E	155·5					
Mice							
		AP_1	Ap_1	aP_1	ap_1		
$AaP_1p_1 \times aap_1p_1$	O	937	960	961	867	6·911	0·075
	E	933·5					
		AB	Ab	aB	ab		
AaBb × aabb	O	935	971	907	921	2·428	0·495
	E	933·5					
		BP_1	Bp_1	bP_1	pb_1		
$BbP_1p_1 \times bbp_1p_1$	O	942	900	956	936	1·829	0·612
	E	933·5					

4. Matings of the type **AaBb** × **AaBb**. Expected ratio of 9 : 3 : 3 : 1.

TABLE IV

Mating		Phenotypes				χ^2	P
Rats							
		AHy	Ahy	aHy	ahy		
AaHyhy × AaHyhy	O	309	94	93	37	1·526	0·681
	E	299·81	99·94	99·94	33·31		
		HHy	Hhy	hHy	hhy		
HhHyhy × HhHyhy	O	377	135	128	34	2·186	0·539
	E	379·12	126·38	126·38	42·12		
		RHy	Rhy	rHy	rhy		
RrHyhy × RrHyhy	O	66	17	20	6	1·044	0·791
	E	61·31	20·44	20·44	6·81		
		DHy	Dhy	dHy	dhy		
DdHyhy × DdHyhy	O	111	42	24	14	5·438	0·145
	E	107·46	35·82	35·82	11·94		
		CHy	Chy	cHy	chy		
CcHyhy × CcHyhy	O	427	136	140	46	0·232	0·972
	E	421·31	140·44	140·44	46·81		
		AD	Ad	aD	ad		
AaDd × AaDd	O	41	19	16	5	1·456	0·697
	E	45·56	15·19	15·19	5·06		
		RD	Rd	rD	rd		
RrDd × RrDd	O	195	52	69	18	3·119	0·376
	E	187·83	62·61	62·61	20·87		

5. Matings of the type AaBbCc × aabbcc. Theoretical ratio of 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1.

TABLE V

Mating		Phenotypes								χ^2	P
Rats											
		ADHy	ADhy	AdHy	Adhy	aDHy	aDhy	adHy	adhy		
AaDdHyhy × aaddhyhy	O	72	71	70	62	70	77	58	66	3·714	0·810
	E	68·25									
		HDHy	HDhy	HdHy	Hdhy	hDHy	hDhy	hdHy	hdhy		
HhDdHyhy × hhddhyhy	O	81	85	70	65	61	63	58	63	9·810	0·201
	E	68·25									
		RDHy	RDhy	RdHy	Rdhy	rDHy	rDhy	rdHy	rdhy		
RrDdHyhy × rrddhyhy	O	67	78	69	60	75	70	59	68	4·389	0·733
	E	68·25									
		HRHy	HRhy	HrHy	Hrhy	hRHy	hRhy	hrHy	hrhy		
HhRrHyhy × hhrrhyhy	O	79	80	81	76	64	67	60	66	6·671	0·465
	E	71·625									
		AHHy	AHhy	AhHy	Ahhy	aHHy	aHhy	ahHy	ahhy		
AaHhHyhy × aahhhyhy	O	88	77	62	67	73	80	64	67	7·742	0·357
	E	72·25									
		AHR	AHr	AhR	Ahr	aHR	aHr	ahR	ahr		
AaHhRr × aahhrr	O	83	83	67	63	81	75	66	68	6·510	0·482
	E	73·25									
		AHD	AHd	AhD	Ahd	aHD	aHd	ahD	ahd		
AaHhDd × aahhdd	O	81	75	65	58	89	63	62	68	11·307	0·127
	E	70·125									
		ARHy	ARhy	ArHy	Arhy	aRHy	aRhy	arHy	arhy		
AaRrHyhy × aarrhyhy	O	69	77	79	65	74	70	62	77	3·684	0·813
	E	71·625									
		ARD	ARd	ArD	Ard	aRD	aRd	arD	ard		
AaRrDd × aarrdd	O	68	69	76	64	79	62	69	64	3·613	0·820
	E	68·875									
		HRD	HRd	HrD	Hrd	hRD	hRd	hrD	hrd		
HhRrDd × hhrrdd	O	88	63	78	72	57	66	67	55	12·242	0·094
	E	68·25									
Mice											
		ADB	ADb	AdB	Adb	aDB	aDb	adB	adb		
AaDdBb × aaddbb	O	475	462	460	509	467	494	440	427	10·626	0·157
	E	466·75									

6. Matings of the type **AaBbCcDd** × **aabbccdd**. Theoretical ratio of sixteen classes in equal numbers.

TABLE VI (RATS)

AaRrDdH$_y$h$_y$ × aarrddh$_y$h$_y$			HhRrDdH$_y$h$_y$ × hhrrddh$_y$h$_y$			AaHhDdH$_y$h$_y$ × aahhddh$_y$h$_y$			AaHhRrH$_y$h$_y$ × aahhrrh$_y$h$_y$			AaHhRrDd × aahhrrdd		
Phenotypes	O	E	Phenotypes	O	E	Phenotypes	O	E	Phenotypes	O	E	Phenotypes	O	E
ARDH$_y$	29	34·438	HRDH$_y$	42	34·438	AHDH$_y$	41	34·438	AHRH$_y$	40	36·125	AHRD	41	34·750
ARDh$_y$	39		HRDh$_y$	47		AHDh$_y$	40		AHRh$_y$	42		AHRd	36	
ARdH$_y$	39		HRdH$_y$	34		AHdH$_y$	43		AHrH$_y$	49		AHrD	41	
ArDH$_y$	44		HrDH$_y$	40		AhDH$_y$	32		AhRH$_y$	31		AhRD	29	
ARdh$_y$	30		HRdh$_y$	31		AHdh$_y$	32		AHrh$_y$	34		AHrd	39	
ArDh$_y$	32		HrDh$_y$	39		AhDh$_y$	31		AhRh$_y$	35		AhRd	33	
ArdH$_y$	32		HrdH$_y$	37		AhdH$_y$	28		AhrH$_y$	30		AhrD	35	
Ardh$_y$	32		Hrdh$_y$	35		Ahdh$_y$	30		Ahrh$_y$	31		Ahrd	25	
aRDH$_y$	40		hRDH$_y$	27		aHDH$_y$	41		aHRH$_y$	41		aHRD	50	
aRDh$_y$	39		hRDh$_y$	31		aHDh$_y$	46		aHRh$_y$	39		aHRd	29	
aRdH$_y$	31		hRdH$_y$	36		aHdH$_y$	28		aHrH$_y$	32		aHrD	38	
arDH$_y$	31		hrDH$_y$	35		ahDH$_y$	30		ahRH$_y$	34		ahRD	30	
aRdh$_y$	31		hRdh$_y$	30		aHdh$_y$	34		aHrh$_y$	43		aHrd	34	
arDh$_y$	39		hrDh$_y$	32		ahDh$_y$	32		ahRh$_y$	32		ahRd	34	
ardH$_y$	27		hrdH$_y$	22		ahdH$_y$	30		ahrH$_y$	30		ahrD	32	
ardh$_y$	36		hrdh$_y$	33		ahdh$_y$	33		ahrh$_y$	35		ahrd	30	
χ^2	10·626			15·621			14·459			12·948			16·086	
P	0·778			0·408			0·492			0·606			0·377	

7. Matings of the type **AaBbCcDdEe × aabbccddee**. Theoretical ratio of thirty-two classes in equal numbers.

TABLE VII (RATS)

AaHhRrDdH$_y$h$_y$ × aahhrrddh$_y$h$_y$		
Phenotypes	O	E
AHRDH$_y$	17	17·219
AHRDh$_y$	23	
AHRhH$_y$	20	
AHrDH$_y$	24	
AhRDH$_y$	12	
AHRdh$_y$	16	
AHrdH$_y$	23	
AHrDh$_y$	17	
AhrDH$_y$	20	
AhRDh$_y$	16	
AhRdH$_y$	19	
AHrdh$_y$	16	
AhrdH$_y$	9	
AhRdh$_y$	14	
AhrDh$_y$	15	
Ahrdh$_y$	16	
aHRDH$_y$	25	
aHRDh$_y$	24	
aHRdH$_y$	14	
aHrDH$_y$	16	
ahRDH$_y$	15	
aHRdh$_y$	15	
aHrdH$_y$	14	
aHrDh$_y$	22	
ahrDH$_y$	15	
ahRDh$_y$	15	
ahRdH$_y$	17	
aHrdh$_y$	19	
ahrdH$_y$	13	
ahrDh$_y$	17	
ahRdh$_y$	16	
ahrdh$_y$	17	
χ^2	25·64	
P	0·728	

8. When albinism is involved, as in a mating of heterozygous coloured self (**CcHh** × **CcHh**), any animal which is albino (**cc**) will not show self or hooded though genetically present. The same is true for agouti and non-agouti. The theoretical ratio is 9 coloured self : 3 coloured hooded : 4 albino.

TABLE VIII (RATS)

Mating	Phenotypes				χ^2	*P*
		CH	Ch	c		
CcHh × CcHh	O	358	127	162	0·364	0·835
	E	363·94	121·31	161·75		
		CA	Ca	c		
CcAa × CcAa	O	362	123	162	0·034	0·983
	E	363·94	121·31	161·75		

9. Matings of rats heterozygous for colour, agouti, and self (**CcAaHh** × **CcAaHh**) will give a theoretical ratio of 27 coloured, agouti, self : 9 coloured, agouti, hooded : 9 coloured, non-agouti, self : 3 coloured, non-agouti, hooded : 16 albino.

TABLE IX

Mating	Phenotypes						χ^2	*P*
		CAH	CAh	CaH	Cah	c		
CcAaHh × CcAaHh	O	278	95	79	39	161	4·263	0·375
	E	275·06	91·69	91·69	30·56	163		

10. The theoretical ratio among progeny from matings of rats heterozygous for colour, agouti, and hair (**CcAaH$_y$h$_y$** × **CcAaH$_y$h$_y$**) or heterozygous for colour, self and hair (**CcHhH$_y$h$_y$** × **CcHhH$_y$h$_y$**) is 27 : 9 : 9 : 3 : 12 : 4.

TABLE X

Mating	Phenotypes							χ^2	*P*
		CAH$_y$	CAh$_y$	CaH$_y$	Cah$_y$	cH$_y$	ch$_y$		
CcAaH$_y$h$_y$ × CcAaH$_y$h$_y$	O	252	78	79	29	104	38	0·867	0·971
	E	244·69	81·56	81·56	27·19	108·75	36·25		
		CHH$_y$	CHh$_y$	ChH$_y$	Chh$_y$	cH$_y$	ch$_y$		
CcHhH$_y$h$_y$ × CcHhH$_y$h$_y$	O	263	93	102	26	114	43	2·935	0·710
	E	270·42	90·14	90·14	30·05	120·19	40·06		

11. A mating of heterozygous coloured, agouti, self, haired rats (**CcAaHhH$_y$h$_y$** × **CcAaHhH$_y$h$_y$**) would give a theoretical ratio of 81 coloured, agouti, self, haired : 27 coloured, agouti, self, hairless : 27 coloured, non-agouti, self, haired : 27 coloured, agouti, hooded, haired : 9 coloured, agouti, hooded, hairless : 9 coloured, non-agouti, self, hairless : 9 coloured, non-agouti, hooded, haired : 3 coloured, non-agouti, hooded, hairless : 48 albino, haired : 16 albino, hairless.

TABLE XI

CcAaHhHyhy × CcAaHhHyhy		
Phenotypes	O	E
CAHHy	180	183·5
CAHhy	60	61·2
CAhHy	63	61·2
CaHHy	50	61·2
CAhhy	18	20·4
CaHhy	24	20·4
CahHy	29	20·4
Cahhy	5	6·8
cHy	104	108·8
chy	38	36·2
χ^2	7·612	
P	0·574	

Among the sixty-five ratios given in Tables I–XI, five depart significantly from the theoretical expectation. D/E for the ratio obtained from **Hh** × **hh** is 4·90, for **Dd** × **dd**, 3·92 and for **P$_2$p$_2$** × **p$_2$p$_2$**, 3·28 (Table I). For **Hh H$_y$h$_y$** × **hhh$_y$h$_y$**, P is 0·027, and for **Hh Dd** × **hhdd**, P is 0·005 (Table III). Four of these five ratios involve the genes for hooded and dilution. Wherever these genes appear in backcrosses, a deficiency in the phenotypes showing these genes always occurs.

Among thirteen crosses giving a theoretical ratio of 1 : 1, ten have the recessive class smaller than the expected, but only three significantly so. The sum of all these monohybrid backcrosses is 9588 dominants : 9153 recessives D/E = 4·71. For the crosses of the type **Aa** × **Aa** and which give a theoretical ratio of 3 : 1, a total of 5070 dominants : 1677 recessives was obtained. In this case D/E = 0·407, which is a very close fit.

For the ratios with four classes expected in equal number, P = 0·0076, and for the crosses giving a theoretical ratio of 9 : 3 : 3 : 1, P = 0·804. For ratios with eight equal classes, P = 0·0014, and for sixteen classes, P = 0·0278.

When the differences between the observed and theoretical classes are all (or nearly all) in the same direction, the larger the number of such divergent cases which are added together the greater will be the significance of the departure from the theoretical expectation.

NOTE ON THE PRECEDING ANALYSIS OF MENDELIAN SEGREGATIONS

By J. B. S. HALDANE, F.R.S.

THE data summarized in Table VII are, I think, unique. However the authors' analysis of them can, I believe, be slightly improved. There are thirty-two classes whose expectation is equal, giving thirty-one degrees of freedom. Now each of these can be specified, and the appropriate value of χ^2 calculated. This is done in my Table I. The first five degrees of freedom are the segregations for single pairs of genes. Thus there were 277 **A** and 274 **a** rats. Hence

$$\chi^2 = \frac{(277-274)^2}{551} = 0{\cdot}0163.$$

The next ten correspond to potential linkages. Thus the degree denoted as **AH** is due to the dichotomy of the total into 281 **AH** and **ah** rats, and 270 **Ah** and **aH**, giving $\chi^2 = 0{\cdot}2196$. Ten more degrees are obtained by considering the associations of gene pairs three at a time. Thus there were 270 (**AHR** + **Ahr** + **aHr** + **ahR**) rats and 281 (**aHR** + **AhR** + **AHr** + **ahr**), giving $\chi^2 = 0{\cdot}2196$. This degree is denoted by **AHR**. Five degrees are obtained by considering four pairs at a time. Thus the degree **AHRD** is given by 274 (**AHRD** + **AHrd** + **AhRd** + **AhrD** + **aHRd** + **aHrD** + **ahRD** + **ahrd**) and 277 (**aHRD** + **AhRD** + **AHrD** + **AHRd** + **Ahrd** + **aHrd** + **ahRd** + **ahrD**) rats, giving $\chi^2 = 0{\cdot}0163$. Finally the degree **AHRDHy** is given by

278 (**AHRDHy** + **AHRdhy** + **AHrDhy** + ... + **Ahrdhy** + ...)

and 273 (**aHRDHy** + ... + **ahrDHy** + ... + **ahrdhy**), giving $\chi^2 = 0{\cdot}0454$.

In each case the principle is the same. We consider the genes 2, 3, 4 or 5 at a time, and divide the rats into two groups, one containing an even number of dominant genes, the other an odd.

Biologically the first five degrees of freedom represent differences of viability between dominants and recessives.

The next ten could represent linkages. However, as there is good reason to think that the five genes concerned are in different chromosomes, deviations greater than could be accounted for by sampling would probably be due to the fact that the effects of the different genes on viability were not additive. For example, Roberts, Dawson and Madden's Table I leaves little doubt that **h** and **d** have an adverse effect on viability, presumably before birth. We further notice that the degree of freedom **HD** has a large, though not significant, χ^2, due to an excess of **HD** and **hd** rats. The actual numbers are 168 **HD**, 137 **Hd**, 125 **hD**, 121 **hd**. If the viabilities were as $1 : 1-\alpha : 1-\beta : 1-\alpha-\beta$ we should expect, on the basis of the single-factor segregations, to find 161·25 **HD**, 143·75 **Hd**,

TABLE I

Analysis of χ^2 in Roberts et al., Table VII

Degree of freedom	Difference	X	χ^2	P
A	+ 3	+0·1278	0·0163	0·90
H	+59	+2·5135	6·3172	0·012
R	+ 5	+0·2130	0·0454	0·83
D	+35	+1·4911	2·2232	0·14
Hy	− 5	−0·2130	0·0454	0·83
AH	+11	+0·4686	0·2196	0·64
AR	−11	−0·4686	0·2196	0·64
AD	−13	−0·5538	0·3067	0·58
AHy	+27	+1·1502	1·3158	0·25
HR	+ 1	+0·0426	0·0018	0·97
HD	+27	+1·1502	1·3158	0·25
HHy	+ 7	+0·2982	0·0889	0·77
RD	− 3	−0·1278	0·0163	0·90
RHy	+ 5	+0·2130	0·0454	0·83
DHy	− 5	−0·2130	0·0454	0·83
AHR	−11	−0·4686	0·2196	0·64
AHD	−25	−1·0650	1·1343	0·29
AHHy	+19	+0·8094	0·6552	0·42
ARD	−23	−0·9798	0·9419	0·33
ARHy	−31	−1·3206	1·7459	0·19
ADHy	− 9	−0·3834	0·1470	0·70
HRD	37	+1·5762	2·4664	0·12
HRHy	−15	+0·6290	0·4083	0·53
HDHy	−13	−0·5538	0·3067	0·58
RDHy	−31	−1·3206	1·7459	0·19
AHRD	− 3	−0·1278	0·0163	0·90
ARDHy	−31	−1·3206	1·7459	0·19
AHDHy	−13	−0·5538	0·3067	0·58
AHRHy	−23	−0·9798	0·9419	0·33
HRDHy	+17	+0·7242	0·5245	0·47
AHRDHy	+ 5	+0·2130	0·0454	0·83
Total 31	− 7	−0·2130	25·639	0·76

131·75 **hD**, 114·25 **hd**, which would give $\chi^2 = 0$ for the degree of freedom **HD**. If the viabilities were as $1 : 1-\alpha : 1-\beta : (1-\alpha)(1-\beta)$, which is perhaps a more plausible hypothesis, we should expect 162·19 **HD**, 143·81 **Hd**, 131·81 **hD**, 114·19 **hd**, giving $\chi^2 = 0{\cdot}001$. So the two hypotheses are indistinguishable except in enormous samples.

The remaining degrees of freedom represent similar biological possibilities as to the non-additive character of the effects of various genes on differential viability. Fisher (1925), who was the first to point out the method here employed for the analysis of χ^2, states that these degrees have "no simple biological meaning". However, the positive value for the degree **HRD** could be mainly due, for example, to the fact that **hRD** rats, of which there were only 58 as against an expectation of 68·875, are more inviable than was to be expected from the effects

of the genes one at a time, which would give an expectation of 66. Such interactions have, of course, been observed by Gonsalez (1923), Timofeeff-Ressovsky (1934) and others.

In my Table I the values of P in the fifth column are read off from those of χ in the third by means of a table of the probability integral, except the last entry, which is calculated by Wilson and Hilferty's (1931) theorem. There are no suspiciously high values of P, and only one (for the degree **H**) which is significantly low. Actually a glance at Roberts, Dawson and Madden's Table I shows that had their Table VII been based on a larger sample the values for **H** and **D** would almost certainly have been significantly low.

In the earlier tables some rats are included which do not figure in Tables VII and XI. Hence for a full discussion of the data each table would have to be analysed separately.

We can now criticize Roberts, Dawson and Madden's final analysis of the data. They point out that the total χ^2 for segregations involving eight equal classes, i.e. three genes at a time, is unduly high. We can see why this is so. Supposing we extracted from Table VII the data on segregations of three genes at a time we should obtain ten tables with seventy degrees of freedom. If these were analysed into their components we should find that we had counted the ten degrees of which **AHR** is typical once each, those of which **AH** is typical three times each, and those of which **A** is typical (i.e. single gene-pair segregations) six times each. We have thus given a quite undue weight to those degrees of freedom which actually show the greatest deviations. Actually each determination of χ^2 by the authors involves one degree of freedom not dealt with in the earlier tables, while the remainder refer to segregations already considered, and often on larger samples, in their earlier tables.

Thus, to take an example, the rat mating **AaHhDd** × **aahhdd** of Table V has a χ^2 of 11·31 for seven degrees of freedom. But the only new information in Table V relates to the difference between **AHD** + **Ahd** + **aHd** + **ahD** and **aHD** + **AhD** + **AHd** + **ahd**, giving $\chi^2 = 1{\cdot}941$ for one degree of freedom. All the other information is already given in Table III, along with some more, for example, concerning rats segregating for **A** and **D**, but not for **H**. The correct method of collating the data is therefore to add up the values of χ^2 for the last degree of freedom in each case.

The expression for the last degree of freedom in a F_2 mating such as **AaDd** × **AaDd** can easily be shown to be $\dfrac{(\mathbf{AD} - 3\mathbf{Ad} - 3\mathbf{aD} + 9\mathbf{ad})^2}{9(\mathbf{AD} + \mathbf{Ad} + \mathbf{aD} + \mathbf{ad})}$. This expression has $\beta_1 = \frac{16}{9n}$, and a nearly normal distribution. Similar expressions can be written down for larger F_2's. Where the mating involves epistacy the formulae for the last degree of freedom are similar. For example, the first mating of Table VIII involves two degrees of freedom, the degree **C** for the segregation of **C** and **c** rats and the degree **H**(**C**) for the segregation of **H** and **h** among **C** rats.

TABLE II

Values of χ^2 for single degrees of freedom, all data

Animal	Degree of freedom	χ^2
Rat	A	2·456
,,	H	10·816
,,	R	0·353
,,	C	0·191
,,	Hy	0·427
,,	D	0·947
Mouse	A	1·630
,,	P_1	1·024
,,	B	0·670
,,	P_2	0·901
,,	Y	0·032
Rabbit	E	0·177
,,	A	0·889
Rat	AA	1·138
,,	HH	1·367
,,	RR	0·016
,,	CC	0·177
,,	HyHy	0·000
,,	DD	0·020
Mouse	P_2P_2	1·255
Rabbit	EE	1·426
1 factor	21	31·012
Rat	AHy	0·629
,,	HHy	0·527
,,	RHy	0·061
,,	DHy	0·323
,,	CHy	0·296
,,	AD	0·175
,,	HD	2·532
,,	RD	1·471
,,	AR	0·001
,,	HR	0·387
,,	AH	0·778
Mouse	AP_1	4·252
,,	AB	0·130
,,	BP_1	0·130
Rat	A_2Hy_2	1·368
,,	H_2Hy_2	1·852
,,	R_2Hy_2	0·081
,,	D_2Hy_2	0·885
,,	C_2Hy_2	0·025
,,	A_2D_2	0·657
,,	R_2D_2	0·012

Animal	Degree of freedom	χ^2
2 factors	21	16·572
Rat	ADHy	0·118
,,	HDHy	0·264
,,	RDHy	2·117
,,	HRHy	0·141
,,	AHHy	0·692
,,	AHR	0·246
,,	AHD	1·941
,,	ARHy	2·934
,,	ARD	1·134
,,	HRD	2·930
Mouse	ADB	2·786
3 factors	11	15·303
Rat	ARDHy	1·746
,,	HRDHy	0·525
,,	AHDHy	0·307
,,	AHRHy	0·692
,,	AHRD	0·065
4 factors	5	3·335
Rat	AHRDHy	0·045
Rat	$H_2(C)$	0·302
,,	$A_2(C)$	0·090
,,	$A_2H_2(C)$	2·530
,,	$A_2Hy_2(C)$	0·980
,,	$H_2Hy_2(C)$	1·700
,,	$A_2H_2Hy_2(C)$	2·578
Epistacy	6	8·180
Grand total	65	75·35

This would be affected by linkage. Its $\chi^2 = \dfrac{4(\mathrm{CH}-3\mathrm{Ch})^2}{9(\mathrm{CH}+\mathrm{Ch}+\mathrm{c})}$. Similarly, for Table IX the last degree of freedom involving the association of A and H among C animals is

$$\chi^2 = \frac{4(\mathrm{CAH}-3\mathrm{CAH}-3\mathrm{CaH}+9\mathrm{Cah})^2}{27(\mathrm{CAH}+\mathrm{CAh}+\mathrm{CaH}+\mathrm{Cah}+\mathrm{c})}.$$

The last component degrees of the sixty-five segregations of Tables I–XI are given in my Table II. Whereas Roberts, Dawson and Madden, for the reason given above, find a significantly large total χ^2, I do not. The probability of $\chi^2 = 75{\cdot}44$ for sixty-five degrees of freedom is $P = 0{\cdot}18$. The whole excess is due to the single-factor segregations, and of these, one gives a significantly large χ^2, while another probably does so. Of the forty-four degrees of freedom involving more than one factor the largest is 4·252, and one value as large as this is to be expected. The order of the degrees is taken from Roberts, Dawson and Madden. If we include the eight cases involving epistacy with the others involving the same number of factors (gene pairs) we find the results summarized in my Table III.

TABLE III

Summary of χ^2 values

Number of factors segregating	Number of degrees of freedom	Total χ^2
1	21	31·912
2	23	16·942
3	14	20·535
4	6	5·913
5	1	0·045
Total	65	75·35

It will be seen that, except for the single-factor ratios, there is no evidence of deviation from Mendelian expectations, and on the whole the fit is very much better than would appear from the analysis of Roberts, Dawson and Madden.

REFERENCES

FISHER, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh.

GONSALEZ, B. (1923). "Experimental studies on the duration of life. VIII. The influence upon duration of life of certain mutant genes of *Drosophila melanogaster*." *Amer. Nat.* **57**, 289.

TIMOFEEFF-RESSOVSKY, N. W. (1934). "Über die Vitalität Genmutationen und ihrer Kombinationen bei *Drosophila funebris* und ihre Abhängigkeit vom 'genotypischen' und vom äusseren Milieu." *Z. indukt. Abstamm.- u. VererbLehre*, **66**.

WILSON, E. B. & HILFERTY, M. M. (1931). "The distribution of Chi-square." *Proc. Nat. Acad. Sci., Wash.*, **17**, 684.

THE USE OF STATISTICAL METHODS IN THE INVESTIGATION OF PROBLEMS OF CLASSIFICATION IN ANTHROPOLOGY

PART I. THE GENERAL NATURE OF THE MATERIAL AND THE FORM OF INTRARACIAL DISTRIBUTIONS OF METRICAL CHARACTERS

By G. M. MORANT

Every naturalist who has had the misfortune to undertake the description of a group of highly varying organisms, has encountered cases (I speak after experience) precisely like that of man; and if of a cautious disposition, he will end by uniting all the forms which graduate into each other, under a single species; for he will say to himself that he has no right to give names to objects which he cannot define.

CHARLES DARWIN, *The Descent of Man*, 1871.

1. INTRODUCTION

THE main aim of physical anthropology is to unravel the course of human evolution, and it may be taken for granted to-day that the proper study of the natural history of man is concerned essentially with the mode and the path of his descent. The general method employed is to compare the physical characters of suitably chosen groups of individuals, and to discover the interrelationships of these groups from an interpretation of the differences found between them. Difficulties are encountered at the outset owing to the nature of the continuum of which the component parts have to be compared. The groups which are most suitable for the purpose in view can easily be recognized in a general way, but it is difficult to define them with precision. There is a lamentable lack of agreement among anthropologists to-day regarding the way in which populations suitable for the purpose of investigating mass descent can best be discriminated. Darwin stressed the difficulty of the problem, and it would be idle to hope that any simple solution of it has been overlooked.

This paper provides a discussion of the statistical approach to anthropological taxonomy. The general thesis maintained in it is that the comparatively new method can be used in a systematic way to discriminate suitable groups, to reveal their interrelationships, and hence to disclose the course of racial history in so far as adequate evidence is available.

It is now more than forty years since Karl Pearson first applied to racial

material the statistical methods which he established and greatly extended.* Recognition of the value of his procedure was slowly won, but the methods are almost universally accepted to-day. It may even be said now that the advantages of their use in physical anthropology are generally taken for granted. There is no longer a need to emphasize the value of measurements, or to point out repeatedly the futility of attempting to derive any useful conclusions from evidence as scanty as that on which many earlier theories were based.

Karl Pearson's teaching in this field has been most widely accepted in so far as methods of reducing group data are concerned, but it is one thing to describe and another to interpret in anthropological terms the situation observed. With regard to such interpretation, he indicated the lines along which he expected that progress would be most profitable, but he did not codify a system or lay down any rigid rules for the guidance of those who wished to follow him. His ideas with regard to this matter could only be grasped by observation of the ways in which he treated particular sets of material. Meanwhile, different anthropologists were drawing deductions from statistical evidence in a variety of ways, some of which are entirely at variance with what came to be known as "biometric" practice, and many of which are irreconcilable *inter se*. The position became chaotic, and it is still in this state.

The view advocated in this paper is that the ways in which statistical methods may be used to supply valid and useful anthropological conclusions can only be determined from wide application of these methods to suitable material. No *a priori* considerations are likely to be of much help here: the nature of the situation has to be examined thoroughly before it is possible to decide on the best ways of treating it to give results of the kind required. The short history of the subject bears clear witness to the fact that the empirical test is always the crucial one. The nature of anthropological groups has now been sufficiently explored, and certain methods of interpretation have now been sufficiently applied, to make possible a just assessment of the value of a particular procedure due primarily to Karl Pearson. He repeatedly stressed the need for adapting statistical theory to practice, and we may follow him in deciding that the value of any new descriptive methods, or modes of interpretation, suggested must be judged from a sufficiently wide application of them.

The statistical nature of anthropological groups will be discussed first, and certain simple but important generalizations which are often ignored by

* The first general treatment of the topic was given in a paper by Cicely D. Fawcett and others (1902) which Karl Pearson edited and arranged. He had previously applied statistical methods to anthropological material in *The Chances of Death and other Studies in Evolution* (1897) and in a series of papers in the *Proceedings* and *Transactions* of the Royal Society of London. Quetelet (chiefly in 1835 and 1871), Galton (in various publications), Stieda (1883) and Witt (1879) had previously used statistical methods in discussing anthropological problems, but they were either not concerned with problems of racial differentiation or else they considered such problems only in a cursory way.

anthropologists can be formulated at this stage. In the second place, different methods of reduction will be considered, and, lastly, the anthropological conclusions which may legitimately be derived from the data reduced in that way will be discussed. In this Part I of the whole paper topics discussed concern the selection of anthropological samples (§ 2), general considerations regarding the treatment of the samples chosen (§ 3), and the forms of intraracial distributions (§ 4).

2. The selection of anthropological samples

Anthropology has been defined as the study of groups, and this definition is appreciated at once by the statistician, since his methods are essentially designed for the treatment of group data. In a consideration of anthropological material the nomenclature of the statistician may be used with advantage: the nature of the processes of description and analysis is thereby made clearer and certain ambiguous biological terms are avoided.

A *population* is defined to be any assemblage of individuals considered and treated as a single group. It may be large or small, and there may be little or much justification for treating it as a single group. The term is a general, but unambiguous, one which can be conveniently used in practice. A *sample* is made up by a number of individuals selected from a population, and the selection is said to be *random* when it is believed that the population as a whole is fairly represented, so that there is no bias favouring any special section of it. A sample of individuals may be said to form a *series*, and these two terms can often be interchanged without loss of perspicuity.

The general method of the physical anthropologist in dealing with new material is: (*a*) to select a sample at random from a particular population of a suitable kind, (*b*) to describe the characters of the individuals comprising the sample, (*c*) to infer from these observations, with greater or less accuracy according to the size of the sample, certain characteristics of the population sampled, (*d*) to make comparisons between the evidence so obtained and that available for other populations described in the same way, and (*e*) to deduce from these comparisons the biological relationships of the new population. It is necessary to make clear distinctions between these successive processes, and failure to do so has sometimes led to confusion.

The description in statistical terms of the first process, i.e. that of sampling, is perfectly precise providing that there is some means of ensuring that the samples are taken at random, but in applying the process to his material the anthropologist is faced with another difficulty. How is he to distinguish the populations from which suitable samples may be taken? In general very little consideration has been given to this question. In comparisons made between samples selected in different ways and representing different kinds of populations, differences between the kinds of sources from which they were derived have

generally been ignored. But in fact such differences are of vital importance, and half the difficulty of the whole process of the statistical analysis of anthropological material is overcome if an effective way of dealing with the problem of selecting suitable samples can be devised. The contention made here is that owing to diversity in the modes of selection of the samples commonly used by anthropologists, *there is often little justification for applying statistical methods* to them in a rigid way.

For the anthropologist the ideal population would be one made up by a number of individuals having a common descent. In the most favourable circumstances such communities cannot be distinguished at all exactly, however, and in the initial stage of his enquiry the anthropologist is generally quite unable to delimit with any approach to precision groups defined in such an abstract way. In practice far less stringent conditions have to be accepted. It may be said that a community made up by a number of individuals whose ancestors—or the majority of them, at least—are believed to have intermarried for a considerable number of generations will be a suitable one to accept as a unit group from which a sample may be taken. Such groups have to be chosen as carefully as possible, having regard to any relevant evidence available. But populations of very different sizes will satisfy the condition stated. The total population of a province and also a small parochial community forming a special part of this total might be considered, and it is possible that these are different in nature and that different conclusions would be derived from them. A further condition, then, which should be satisfied wherever possible, is that the population considered is a large one of the regional rather than of the parochial kind. Experience, discussed below, has shown that large communities are always more suitable for consideration than small ones when the purpose in view is the analysis of more remote origins. But if the only sample which can be obtained does represent a population of the parochial kind—as may be the case if a series of skeletons from a single cemetery is the only material available—then it should be fully recognized that this may have been a special part of a larger population which might more profitably be considered as a unit group, and allowance should be made for the peculiarity of the source.

Having selected a number of samples from what appear to be suitable populations, the anthropologist may compare their statistical features, such as the forms of distributions, measures of variation and correlations they supply. If the majority of the samples are found to possess certain particular characteristics, while departures from the rule are occasionally met, then it may be legitimate to conclude that the peculiarity of the exceptional cases is due to the fact that the populations they represent are unsuitable for the purpose in view. The empirical investigation of samples known to represent unsuitable populations can aid examination of the matter. The practical procedure which appears to be most profitable is this: samples believed to be of the right kind are selected

and then certain tests, derived from experience of the statistical characteristics customarily exhibited by such samples, are applied in order that abnormal ones may be revealed and rejected as unsuitable for further use.

The initial process of selecting samples must be considered rather more fully. In practice it may be advisable to take into account information of several kinds. The most important factor is usually geographical position, but the separation of series representing different social classes may be desirable, and archaeological, historical or linguistic evidence may suggest that the partitioning should be carried out in a particular way. In dealing with skeletal material the time factor may be an important one to consider; it may be well to treat a pagan and a Christian Anglo-Saxon series separately, for example, and not to assume that the two represent indistinguishable populations. The general rule is that the total series for which data are available should be split up in a natural way into as many subseries representing distinct populations of a considerable size as possible, providing that each subseries selected is large enough to give comparisons of value. The way in which this can best be done depends on circumstances which are peculiar to each set of material. The anthropologist is thus obliged to plan his survey of a particular region with evidence of several different kinds in view. The majority of these do not relate to physical characters, but he has a perfect right to consider, as relevant to his biological enquiry, any evidence which gives some indication of the subgroups of the total population which are such that intermarriage normally takes place, or took place, within rather than between them.

The procedure described is necessarily rather vague and of an arbitrary nature: any subdivision adopted initially is experimental. In practice choice is largely controlled by the amount of information available. Suppose, for example, that the total sample for which measurements are available represents people coming from all parts of a particular country. If there are enough individuals they may be divided into subsamples on a regional basis, or the population of each region may be thought of as subdivided into county groups, say, and each of these may be further subdivided into parochial communities, which may be split up into small groups of closely interrelated people. A hierarchy of groups within groups can thus be imagined, but in practice it will not be possible to carry the process of subdivision beyond a certain stage, as further dissection would lead to subsamples too small to yield comparisons of value. There can be no assurance that the divisions based on geographical or other considerations will be the most effective for the purpose in view, but some such divisions have to be adopted in order to disclose the nature of the total population.

At the outset of his survey of new material relating to the population of a particular region, the anthropologist thus requires a knowledge of the size of the smallest series which can be used profitably for the purpose of the classification

in view. This minimum requirement cannot be determined from any *a priori* considerations: the point has to be settled after an experience has been gained from the comparison of series of different sizes. The conclusions reached in this way are discussed below.

However ample his material is, the anthropologist engaged in investigating remoter origins and relationships should choose his unit groups, *whenever possible*, in such a way that they relate to populations of a considerable size. Family and parochial groups should certainly be considered too small. The need for restricting comparisons to larger groups than these if possible is evident on account of both *a priori* and *a posteriori* considerations, but it is frequently overlooked.* It should be anticipated that series representing small communities are likely to be biased, not random, samples of regional populations. The existence of local variants of a widespread population derived from a single source has to be recognized, and hence special precautions have to be taken in drawing deductions from samples representing local groups. In conformity with these expectations, it is commonly found in treating a total group of the national kind that samples representing small subgroups of it tend to differ more than samples representing large subgroups. It is customary to find in such a case that up to a certain point there is greater uniformity in the statistical attributes of the samples—quite apart from fluctuations due to random sampling—according as the groups represented by the samples become larger. Greater diversity has to be expected when smaller subgroups are compared. If the anthropologist concerned with the broader taxonomic problems of his subject treats groups of the family or parochial kind, he will thus be in danger of losing his way among the trees when he should be taking a bird's-eye view of the wood as a whole.

It is unlikely that an example of the statistical conception of a perfectly homogeneous population is ever encountered in dealing with anthropological material. The kind of continuum which has to be analysed by making comparisons between its component parts is of a peculiar nature. It would be an advantage if the subsections of it considered for the purpose of racial classification were always populations of a considerable size, and for many sets of anthropometric

* The writer had overlooked it at one time, and he was accordingly admonished in a letter from Karl Pearson dated 17 July 1924, in which the following passage occurs:

"May I give you an analogy? In a little Yorkshire village there are two manors—two small squires, little better than yeomen—and they kept on intermarrying their family members. About one-third of the churchyard contains, for at least two centuries, the graves of these folk. You could get at least 30, and possibly 100, crania from that third of the churchyard which would differ very sensibly from the skulls in the remainder of the yard. It would arise solely from the fact that we are dealing with an inbred population, which possessed certain characters which raised it above, or at least differentiated it from, the remaining population. I don't suppose there was any racial difference between these little squires and their neighbours, they were all ultimately of Danish descent. But simply one or two interbreeding families were buried in one earth. Now my analogy has for its bearing the danger of picking out 20 crania and saying these differ from the rest."

data this condition is satisfied. Not infrequently, however, and particularly in the case of excavated skeletal material, the only samples available for a particular regional group are of the parochial kind.

The Anglo-Saxon skeletons preserved in museums actually come from a number of scattered cemeteries, and there is no long series from a single cemetery. The fact that they were dispersed should be considered a real advantage when the object in view is to determine the characters of Anglo-Saxons in general, or of the larger subsections of this total population. If all the specimens had come from a single cemetery it would have been difficult *not* to assume that the series, particularly if it were a large one, could be accepted *faute de mieux* as representing the total Anglo-Saxon population, and to draw conclusions from a comparative study on this assumption. In fact there is good reason to suspect that the physical characters of the people forming any small Anglo-Saxon community differed appreciably in their averages from those of any large section of the total population.

In practice it is seldom possible to exercise much choice in selecting samples. In many instances the only material available representing a past population of the national kind consists of one or more series of skeletons from one or two cemeteries. The anthropologist cannot be expected to neglect all series which are not ideally suitable, but he should remember that several he is obliged to use are not ideal for the purpose in view. Allowance has to be made for known differences between the samples used dependent on the diverse ways in which they were selected.

3. The statistical treatment of anthropological samples

In treating his problems of classification the anthropologist starts by selecting series of individuals representing different populations which appear to be more or less suitable for the purpose. Data have been recorded for large numbers of such samples relating to living peoples in different parts of the world and to extinct peoples represented by series of skeletons. It is recognized that these samples are taken from groups of different kinds, and that allowance will have to be made for this fact in comparing them. Apart from this diversity which is appreciated, there is no guarantee that all the samples are suitable for the purpose of examining group relationships. Some may be entirely unsuitable because they represent populations of exceptional kinds, and their peculiarity in origin has to be detected by examination of the samples themselves. For example, a certain number may be found to have peculiar characteristics which suggest that they should not be used, while the majority conform to a particular type. In practice it is necessary to appeal to experience derived from the treatment of a large number of series in order to justify the use of certain tests employed to discriminate between suitable and unsuitable material.

A question which might be asked is: What is the nature of anthropological

samples selected in the way described?—what features are common to all of them and in what ways do they differ? The statistician—who is accustomed to thinking of problems in terms of populations and samples—points out at once that, though this question may be of the right kind, yet it is not put in the right way. He reminds the anthropologist that his ultimate concern is with populations, and that the samples are really of interest only in so far as they supply information regarding populations. A moment's consideration shows that this is, or should be, the view of the anthropologist. If he collects information relating to 100 Greenland Eskimos, say, it is only in order to arrive at generalizations regarding either all Greenland Eskimos, or else some section of them which is still a population many times larger than the sample observed. Hence the question should be: What generalizations regarding populations can be deduced from samples selected in the way described?—what features are common to all the populations and in what ways do they differ?

The fact that his ultimate concern is with populations and not with samples of them is not likely to be entirely overlooked by the anthropologist, but there is a real danger that he may assume in particular instances that the characteristics of samples are precisely the same as those of the populations they represent. Statistical treatment has a great advantage in this connexion, since it keeps the distinction referred to continually in view and provides a systematic and—as far as circumstances permit—precise method of reaching the generalizations required.

In this paper the only kind of information regarding samples which will be considered is that provided by measurements, whether of series of living people or skeletons. The general problem can be viewed in the same way if non-metrical characters are dealt with, but a different statistical treatment is required for them. All the individuals referred to will be supposed adult, and for such the measurements can be supposed, in general, to be unaffected by the age of the individual. The data for males and females have to be considered separately, but the generalizations reached are the same for the two sexes. Data relating to numerous series are available for a particular set of characters commonly recorded by anthropologists: nearly all of these measurements concern either absolute size (chords and arcs) or shape (indices and angles). In the case of the skull, for example—the data for it being more abundant than those for all other bones of the skeleton put together—the measurements are designed to give a description of the size and shape of the skeleton of the head considered as a whole and of all its principal parts. The majority of the anthropometric data for living people commonly recorded provide indirect measures of the size and proportions of different parts of the skeleton.

With the object of obtaining information regarding the distribution of a particular character in the population represented, the following features of the distribution provided by a sample will be considered:

(i) its form,

(ii) its scatter, that is to say the variation exhibited by the readings for the individuals composing the sample,

(iii) its central tendency, that is to say the value of the character (or *variate*) about which the readings can most conveniently be considered to be scattered.

These three features of the distributions are first considered in the case of characters dealt with singly. It is also necessary to gain information regarding the ways in which different characters are associated in individuals, and hence it is necessary to examine:

(iv) the correlations of different pairs or groups of characters in individuals.

A sample of the kind so far considered is made up by a number of individuals believed to belong to a population such that its members have been chiefly intermarrying with one another for a number of generations. This is called an *intraracial* sample—a rather unsatisfactory term to use at this stage since the concept of race has yet to be defined. An examination of the forms of the distributions and of their variabilities in the case of metrical characters considered singly, and for a considerable number of intraracial samples, shows that the only comparisons of them likely to give results of any anthropological interest must be of a certain kind defined in later sections of this paper. Statistical treatments of other kinds are seen to be unprofitable. The same evidence makes it clear that the most appropriate measure of the central tendency of any one of the distributions to use is the arithmetic mean, or average.

Attention is thus focused on mean measurements, and different ways of treating them have to be considered. The conception is introduced of samples for which the units are not actual individuals—as for intraracial samples—but abstract beings such that each has metrical characters equal to the averages for the particular group which it represents. This idea of *l'homme moyen* is not new in anthropology: it was clearly enunciated more than 100 years ago by Quetelet (1835). An assemblage made up by a number of *hommes moyens* representing different populations is called an *interracial* sample, and in the case of a particular character the mean measurements for a number of samples representing different populations will provide an interracial distribution.

In order to gain a more complete knowledge of the statistical nature of anthropological samples, required for the purpose of determining the ways in which they can best be treated in practice, it is necessary to take cognizance of other features of them in addition to those listed above. Interracial considerations are involved in examining:

(v) the ways in which the averages for different characters distinguish certain sets of populations,

(vi) the forms and variabilities of interracial distributions, and

(vii) the interracial correlations of characters.

In the course of the investigation of these topics it is possible to obtain estimates of the minimum sizes of samples which are required, under different

conditions, in order to provide intergroup comparisons of value. One of the advantages of the statistical approach to the problem of classification is that it makes it possible to determine with some precision the least amount of evidence which must be available in order to yield useful conclusions. The history of physical anthropology bears clear witness to the fact that some control of this kind is a vital need.

In the following section of this paper the forms that occur in practice of the intraracial distributions of anthropometric characters are discussed. It is hoped that this will be followed in later parts by discussions of the other topics listed above—viz. other intraracial characteristics of distributions and characteristics of interracial distributions—and, finally, by a discussion of methods of comparing samples.

4. THE FORMS OF INTRARACIAL DISTRIBUTIONS

It can be stated categorically that the distributions of measurements for the vast majority of samples that occur in anthropological practice tend to conform closely to the normal curve. Quetelet's suggestion (1871) that this is so has been confirmed by data relating to numerous series of living people and skeletons from all parts of the world. In general, the samples from a particular population tend to give a closer and closer approximation either to the normal or to a very similar form of continuous curve according as their sizes are increased. Hence it is safe to infer that the characteristic type is unimodal and symmetrical. This generalization applies to all absolute measurements and, in spite of a theoretical qualification mentioned below, also to all measurements of shape (indices and angles) for which the matter has been adequately investigated.

There has been considerable confusion regarding interpretation of the forms of frequency distributions in the discussions of this topic provided by some anthropologists to whom statistical conceptions were unfamiliar. In the early days of biometry the fallacy of drawing conclusions of certain kinds from peculiarities of the curves provided by samples was repeatedly stressed, but this warning is still occasionally ignored. The chief errors made with regard to the matter have been in supposing (*a*) that a small sample—made up by fewer than 100 individuals, say—is capable of giving an adequate estimate of the form of the distribution in the population sampled, and (*b*) that mere inspection of the diagram for a sample is sufficient to reveal the information required. In this connexion numerous examples may be found of the fallacy of supposing that the features of a sample are precisely the same as those of the population from which the sample was taken. The general appreciation by anthropologists of the perfectly clear implications of a simple sampling experiment such as the one to be described would still save much fruitless discussion.

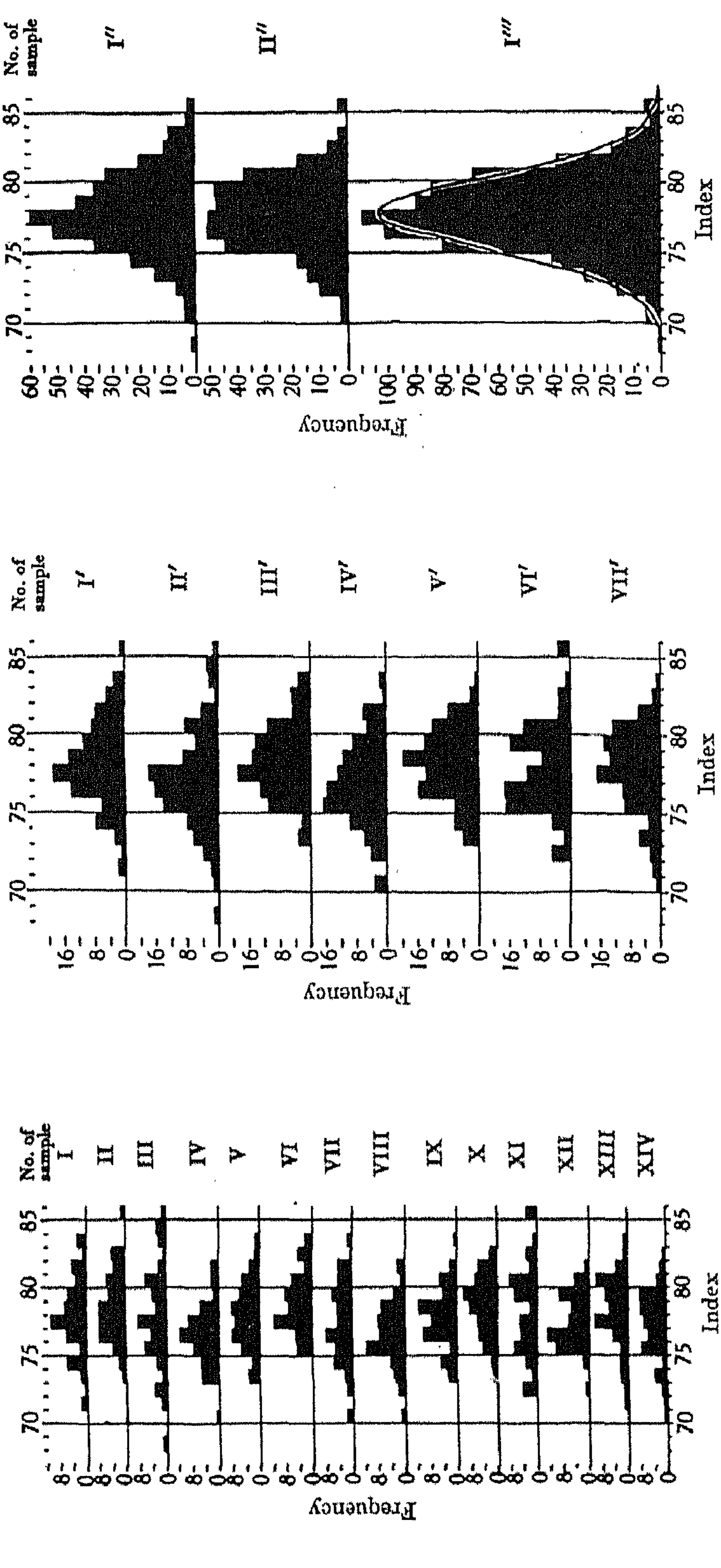
The head lengths and breadths of 1314 soldiers who were natives of Lanarkshire have been given by Tocher (1924). The order in which the first 700 men are arranged in the table appears to have been entirely random as far as the

measurements are concerned.* The cephalic indices were calculated for this sample and Fig. 1A shows the distributions obtained by taking successive groups of 50 from the first (nos. 1–50) to the fourteenth (nos. 651–700). Each one of these subsamples provides an estimate of the distribution of the cephalic index for Lanarkshire men in general, and any one of these subsamples might have been the only one available relating to the total population. The distributions in Fig. 1A clearly exhibit a great diversity of forms. They are all alike in showing a majority of individuals with indices between 75 and 80, and in either covering a continuous range or showing a few outlying values detached from the main body. Otherwise, there is little agreement between them. Some show a more or less gradual rise to a maximum frequency and then a gradual fall (as no. X), while others appear to have two (nos. XII and XIII) or more (nos. III and XI) distinct peaks. The fact that inconsistencies are found in these respects makes it obvious that the series are too small to provide any reliable information as to the existence of such features in the distribution for the population they all represent.

Fig. 1B shows the same 700 measurements treated in successive subsamples of 100, so that the first series (I′) is the sum of the first two (I and II) in Fig. 1A, and so on. The distributions for these larger series show more uniformity—the maximum frequency being for the range 77–78 in four cases out of the seven, for example—but they still differ very appreciably among themselves in some respects, and one (VI′) appears to be clearly of a bimodal form. It is evident that the measurements of more than 100 individuals are required to give any reliable indication of the details of the distribution of the cephalic index in the parent population.

The top and the middle distributions in Fig. 1C relate to the samples made up by the first 350 (I″) and the second 350 (II″) Lanarkshire men, respectively. Both distributions appear to be of symmetrical form, but the first shows an outstanding maximum frequency while the second is "flat-topped" (platykurtic). The conclusion must be that it is unsafe to draw deductions regarding some features of the distribution for the parent population by merely inspecting the distributions of samples, even if they are as large as these two. The bottom diagram in Fig. 1C relates to the total sample of 700 men considered. It appears to be slightly asymmetrical, but comparison with the corresponding areas of the

* After some number a little above 700 the order in which the individuals are arranged is clearly not random. They were evidently taken in successive small groups such that each group had a restricted range of head lengths, and consequently the distributions of the cephalic index for samples of 50 (nos. 701–50, and so on) are very peculiar. Dr Tocher has kindly answered my enquiries with regard to this matter, and he informs me that the departure from randomness was not appreciated when the measurements were taken. It may have been due to the fact that at one depot the officers arranged the privates measured in batches having hats of the same size, in the belief that this would help the anthropologists! Departures from a random selection which may affect the distributions of characters are easily overlooked.



A. Successive samples of 50.

B. Successive samples of 100.

C. Successive samples of 350 (I″ and II″) and the total sample (I‴) fitted with a normal curve.

Fig. 1. Distributions of the cephalic index for subsamples of a total sample of 700 men from Lanarkshire.

normal curve fitted to it shows that the divergences of the histograms from their theoretical values on the hypothesis of normality are all small relative to the sizes of these frequencies.

This example is sufficient to demonstrate that it is entirely unsafe to attach any significance to the peculiarities of the distributions for small samples—any made up by fewer than 200 individuals, say—and that mere inspection of the forms of distributions for samples which would usually be considered adequate for most statistical purposes is liable to mislead. The anthropologist who deduces theories of racial mixture or relationship from the "peaks" of small distributions is deceiving himself by building on a statistical foundation which is unsound owing to the inadequacy of the evidence. He is trying to get more information from his material than it can possibly provide with any assurance of correctness, and forgetting that a sample only provides estimates of the features of the population represented.

It is found in anthropological practice that distributions for small samples frequently appear to be markedly skew, or to be bi- or multimodal, but that those for large samples scarcely ever exhibit any of these peculiarities. The writer is unable to give any examples of a distribution, for the kind of sample now considered, which relates to more than 300 individuals and which fails to show a close approach to the form of a unimodal and symmetrical curve, with the exception of two discussed in the appendix below. The occurrence of more than one mode which appears to be definitely outstanding, or of any appreciable degree of skewness, is apparently never exhibited by the distributions of characters for large samples of the kind encountered in anthropological practice. The conclusion must be that when such peculiarities are found for the distributions of small samples they merely demonstrate that small samples are subject to large "errors" of random sampling. The "peaks" can have no anthropological or genetical significance. Their unstable nature can often be demonstrated by merely grouping the individual measurements in a different way, as this may change the appearance of the distribution to an appreciable extent.

If it can be demonstrated adequately that the distributions which occur in practice almost invariably indicate that the parental distributions must be closely similar in form to the normal curve, then this is obviously a conclusion of great anthropological importance, since it implies that the kind of statistical analysis likely to be at all profitable is severely restricted owing to the nature of the material. Evidence of a more adequate statistical kind regarding the forms of distributions will now be considered.

Values are given below of probabilities (χ^2, P test) obtained by comparing various distributions with the theoretical frequencies obtained by fitting normal curves to them. Those for the cephalic indices of series of Lanarkshire men for samples of 100 or more (see Fig. 1, B and C) are in Table I. The lowest P found is 0·03, indicating that for samples of the size (100) actually drawn at random

from a normally distributed population, 1 in 33 would be expected to give a less good correspondence with the theoretical distribution than that exhibited by the sample in question. All the other P's are much higher, indicating that the estimates provided by the samples bear a much closer resemblance to the normal form in these cases.

TABLE I

Probabilities (χ^2, P test) that the distributions of the cephalic index for different subsamples of 700 Lanarkshire men indicate that the character was normally distributed in the parent population

	Successive samples of 100 individuals*						
	1st	2nd	3rd	4th	5th	6th	7th
P	0·62	0·34	0·80	0·88	0·79	0·03	0·23

	First 350 individuals†	Second 350 individuals†	700 individuals‡
P	0·77	0·64	0·70

* Distribution in 9 groups. † Distribution in 13 groups. ‡ Distribution in 15 groups.

Table II gives the P's for distributions of the cephalic index in the case of 15 series of male skulls of the kinds which normally occur in anthropological practice, each series being made up by more than 100 specimens. These were selected merely because they happened to be the ones of the size required most easily accessible to the writer. The first 11 are believed to be homogeneous in the sense that any subseries of them likely to be chosen for anthropological purposes would not show significant differences in any statistical constants. All but one of the P's for these are high enough to provide excellent support for the hypothesis that the cephalic index was normally distributed in the parent population. The single exception is for an Egyptian series (3) which is known from other evidence to have statistical features which are typical for "homogeneous" samples, so the low P found for the cephalic index may well be attributed to the vagaries of chance selection.

Series 12 and 13 are known to be heterogeneous to some extent, since the means of the cephalic index and some other characters for their subseries show differences which are significant though small. The P's for these two are both high. Series 14 and 15 are known to represent populations which are unsuitable for the purpose of investigating group relationships, since they are exceptionally

variable. This is clearly indicated by the standard deviations for the cephalic index given in the right-hand column of Table II, and also by those for certain

TABLE II

Probabilities (χ^2, P test) that the distributions of cephalic indices provided by various series of male skulls indicate that the character was normally distributed in the parent populations

No. of series	Series*	No. of skulls	Mean	No. of groups	P†	$\sigma \pm$ S.E.
	(*a*) Series believed to be homogeneous					
1	Guanche	245	76·0	11	0·998	2·30 ± 0·10
2	Eskimo (St Lawrence Is.)	156	77·1	10	0·904	2·39 ± 0·14
3	Egyptian (26th–30th dyn.: Gizeh)	866	75·0	16	0·011	2·65 ± 0·06
4	New British	114	72·2	11	0·763	2·74 ± 0·18
5	English (Camb. dissecting rooms)	118	75·8	11	0·157	2·93 ± 0·19
6	Eskimo (Greenland)	190	71·3	13	0·694	3·00 ± 0·15
7	Czech	105	82·9	11	0·470	3·17 ± 0·22
8	English (17th cent.: Whitechapel)	131	74·3	7	0·965	3·20 ± 0·20
9	German (Reihengräber)	220	73·8	13	0·876	3·35 ± 0·16
10	English (17th cent.: Farringdon St.)	135	75·4	11	0·137	3·48 ± 0·21
11	Egyptian (18th–21st dyn.: Thebes)	167	75·0	13	0·283	3·55 ± 0·19
	(*b*) Series showing small but significant differences between component subseries					
12	Egyptian (predynastic: Naqada)	165	72·7	13	0·666	2·80 ± 0·17
13	Swiss (Valais)	453	83·9	19	0·423	4·01 ± 0·13
	(*c*) Heterogeneous series					
14	French (mediaeval and modern)	1000	79·6	22	0·000‡	4·32 ± 0·10
15	English (Bronze Age)	151	78·8	12	0·660	5·42 ± 0·31
	(*d*) Artificial mixture of two series					
2+12	Eskimo + Egyptian	321	74·8	16	0·297	3·43 ± 0·14

* The series are the ones described in the following sources, which give the individual measurements in the majority of cases:

(1) Guanche—Hooton (1925).
(2) Eskimo—Hrdlička (1924).
(3) Egyptian—Pearson & Davin (1924).
(4) New British—Muller (1905).
(5) English—Duckworth (1917).
(6) Eskimo—Furst & Hansen (1915).
(7) Czech—Schiff (1912).
(8) English—Macdonell (1904).
(9) German (pooled)—Morant (1928).
(10) English—Hooke (1926).
(11) Egyptian—Schmidt (1886).
(12) Egyptian—Fawcett (1902).
(13) Swiss—Pittard (1909–10).
(14) French—Topinard (1885, p. 388).
(15) English (pooled)—Morant (1926).

† The P's were found from Table XII in *Tables for Statisticians and Biometricians*, Part I, the n' used being the number of groups, less 2 (i.e. degrees of freedom were taken as number of groups, less 3).

‡ 0·00003.

other characters. One of these heterogeneous series (14) gives a P so low as entirely to disallow for practical purposes the hypothesis that the index was

normally distributed in the parent population, but the P for the other (15) is higher than several of those in section (*a*) of the table. The test thus fails entirely to distinguish between all suitable samples, on the one hand, and all unsuitable ones, on the other. Another example of its failure in this respect is recorded in section (*d*) of the table, the data there relating to an artificial mixture of two series having the very considerable difference of 4·4 units in their mean cephalic indices. The P of 0·3 obtained gives no indication of the strange origin of this sample.

It is easier to point to series which are clearly unsuitable for purposes of classification than to any which would confidently be expected to be entirely satisfactory. A distribution which appears as likely as any to reveal the unsatisfactory nature of a series of the former kind has been provided by Herskovits (1930, p. 151). It relates to the thickness of the lips of 959 male adult "Negroes" in the United States. While a certain proportion of these men are believed to be of pure African origin, the majority represent varying grades of miscegenation of Negroes and members of other ethnic groups, with European ancestors predominating but some admixture of American Indian blood. The character is obviously one which makes a marked distinction between the types of the two major groups of parental populations involved. The distribution for it gives a P of 0·536 (13 groups). Examples of the same kind relating to samples which are obviously unsuitable for purposes of group classification—such as mixed "white" samples in the United States, or artificial mixtures of series which differ markedly in some respects—might be multiplied indefinitely, and it is quite usual in examining such material to find distributions which give high P's when fitted with normal curves. Trevor (1938) has observed this situation for all the characters he examined when dealing with data for living populations in different parts of the world derived from the crossing of European and non-European peoples.

It is actually found that the vast majority of the distributions of measurements that occur in anthropological practice satisfy a test which shows the hypothesis that the character was normally distributed in the parent population to be not at all unacceptable. The high values of P usually found when comparisons are made with normal curves fitted to the data do not demonstrate that the characteristic type of distribution for all but exceptional, and presumably unsuitable, populations is absolutely normal, but they are quite good enough evidence to suggest that the population distributions must be closely similar to unimodal and symmetrical forms.

This generalization applies almost equally well, as far as is known, to all measurements, whether of shape or size, of series of living people or skeletons. Furthermore, it apparently applies equally well to series from groups of any size which are likely to be considered as populations by the anthropologist. Of the series of skulls dealt with in section (*a*) of Table II, some came from single

cemeteries while others were made up by combining specimens from several scattered cemeteries. Samples from small communities of the parochial kind usually provide distributions which approximate to the form of the normal curve neither more nor less closely than do samples of the same size selected at random either from the populations of provinces or from those of countries. The form of the distributions of metrical characters fails entirely to differentiate populations of very different sizes.

The fact that a close approach to normality is almost invariably found suggests that this condition can safely be accepted as one which must be satisfied by the distributions provided by samples before they can be accepted as suitable for group comparisons. It has been seen, however, that the distributions of samples which are known to be unsuitable for purposes of classification may also satisfy the condition. Hence the test must be supposed a necessary but by no means a crucial one. In the case of a particular sample the theoretical test might be that all characters should show reasonably high probabilities that the distributions in the population were normal. In practice, of course, the question can never be examined for more than a small number of characters—very seldom greater than 50—but it is important to appreciate that the test should be applied, in theory at least, to all characters for which data are available. The demonstration that a single distribution indicates a close approach to normality is no evidence that any other distribution for the same sample will do so.

In general, useful anthropological results of a statistical nature can only be derived from the conjoint investigation of data for several characters. Different characters are quite likely to suggest different conclusions, and the kind of information required has to be obtained by taking the evidence of a sufficient number into consideration. To take a hypothetical example, it may be assumed that in the case of the crossing of two populations (A and B) the one derived from them (C) will only be likely to show any departure from the usual form of distribution—such as bimodality—in the case of those characters which show a clear difference between the averages for A and B. But experience shows that the two parental groups are likely to show clear differences in their averages for only a small proportion of the characters compared, even if A and B belong to quite distinct ethnic groups. Hence peculiarity in the distributions for the hybrid population would be expected only in the case of a small proportion of the characters examined, and it is necessary to examine as many as possible in order to ensure that such a peculiarity has not been overlooked. Furthermore, it should be appreciated that the chance of departures from normality being found is very different for different characters. This point is discussed below.

A test which might be applied in practice could specify that the distributions of all metrical characters recorded for a particular sample examined should give P's, when compared with normal curves fitted to them, greater than some

arbitrarily chosen value—0·001, say. If one or more of the distributions were found to give values less than the limit then the sample might be considered so peculiar that it should not be used for purposes of classification. It is known, however, that many unsuitable samples would pass such a test, and hence it is of little practical value. Its value is lessened further by the fact that when a clear departure from normality is found the peculiarity of the distribution can nearly always be detected more easily on account of its abnormally large variation: this is so for series 14 in Table II.

Much labour has been expended in making a detailed analysis of the distributions of characters provided by anthropological samples which are adequate in length for statistical purposes. Significant but slight degrees of skewness have sometimes been recorded. It has not been shown, however, that such examples provide any conclusions which are of the kind needed by anthropologists. Usually there is no guarantee that the samples were chosen absolutely at random, and the slightly peculiar but erratic characteristics sometimes observed may well be due to biased selection.

It has been shown that if the components of an index (A/B) are normally distributed, then the distribution of the index conforms more closely to a Pearson Type IV than to the normal curve. This topic has been discussed by Merrill (1928) and Fieller (1932). A Type IV curve is unimodal and may be almost symmetrical and very similar in form to the normal. In fitting normal curves to a series of distributions it is generally not possible to observe any tendency for indices to give lower values of P than absolute measurements. Elderton & Woo (1932) made a detailed study of the forms of the distributions for a single long series of Egyptian crania in the case of an unusual set of measurements relating to cranial bones considered singly. They concluded: "that the distributions of characters measured on the individual bones of the skull are not of normal type, but rather that the skewness and kurtosis of such distributions are peculiar to the individual measurement." This is a point of considerable theoretical interest, though all the departures from normality observed were in fact slight. The conclusion that the population distributions almost invariably bear a close similarity to the form of the normal curve in the case of all kinds of measurements appears to be sufficiently exact for almost all practical purposes.

This is a result which anthropologists are very loath to accept in spite of the clear evidence of a large amount of material which supports it. The belief that markedly skew and bi- or multimodal distributions must be found was very generally held thirty years ago, and it has not been abandoned yet. The fact that expectations of this kind have been categorically denied to him is a matter of prime importance to the anthropologist, as it suggests that certain methods of analysis are quite impracticable on any valid statistical lines. He still lives in hope of discovering a method of dissecting a sample in such a way as to disclose

the sources from which it was derived. Judging largely from the prevalence of unimodal and symmetrical distributions which offer no scope for dissection, but also to some extent from other characteristics of samples to be considered, the statistician says that it is idle to entertain such a hope in dealing with metrical data. He suggests that statistical tests may profitably be used to distinguish suitable from unsuitable samples, but that those selected for further use must be kept intact and somehow be compared with one another as entities in order that conclusions may be deduced regarding the relationships and descent of the populations represented.

The difference between the two points of view is fundamental. Repetition of the assertion that the vast majority of the samples with which they deal in practice do not indicate either appreciable skewness or more than one mode in the populations represented has had little effect in modifying the methods of anthropologists in general, or even of those who commonly use statistical methods. It is possible to go further, however, and to offer an explanation of what appears at first to be a surprising uniformity in one characteristic of diverse kinds of material.

It is an observed fact that the variabilities of all samples which can be accepted for purposes of classification are very considerable and that they tend to be similar in the case of all populations for which adequate data are available, the earliest of these having existed in Egypt about 5000 B.C. Absolute equality in the populations is not indicated, but intraracial variabilities all lie within a fairly restricted range in the case of any particular character. Evidence in support of this statement will be provided in the following section of this paper.

In the case of a mixture of two populations, a distribution would only be expected to be bimodal if the difference of the means for the two populations happened to be appreciable compared with the two intragroup variabilities. An estimate of the chance that the difference in question is likely to be appreciable will be given by the ratio of interracial to intraracial variabilities, and this is likely to be different for different characters. It has usually been assumed that the differences between the types of populations defined by average measurements are often large compared with the differences between individuals belonging to the same population, but though this situation may be true for skin colour it is not observed in the case of any metrical character for which data are available. With the evidence of skin colour in view, anthropologists have always been inclined to over-estimate the interracial variability and under-estimate the intraracial variability of other characters. The evidence of measurements corrects this tendency and shows that for metrical characters the situation is markedly different from that which has often been postulated.

Examples relating to particular characters will now be considered. Fig. 2 shows the distributions of stature for three series of men. The first relates to a tribe of Congo pygmies (Efé) recorded by Šebesta & Lebzelter (1933): it has a

mean of 1430 mm., which is very close to the smallest given for any people in the world, and a standard deviation of 51·4. The third relates to the Dinka tribe of Nilotic negroes, from unpublished data of A. MacTier Pirrie,* and its mean of 1804 mm. appears to be the largest on record, the standard deviation being 79·7. The second distribution relates to men from a province of Japan recorded by Matsumura (1925), and it has a mean of 1622 mm., which is almost midway between those for the extreme series, and a standard deviation of 53·1.

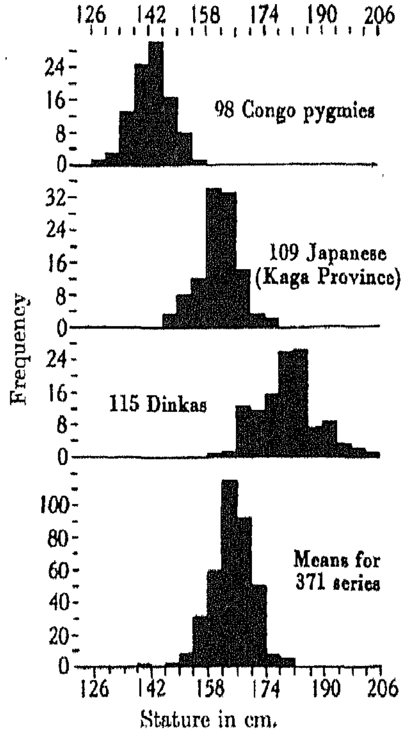


Fig. 2. Intraracial distributions of statures representing an extremely short, an intermediate and an extremely tall group, and an interracial distribution.

There is seen to be an absolute separation between the ranges of the statures for the pygmies and the Dinkas, the tallest member of the former group being shorter than the shortest of the latter, though it is probable that the distributions for larger samples would overlap to some extent. A mixed series made up by taking 100 pygmies and 100 Dinkas would obviously be bimodal. A distribution which would clearly be bimodal would also be provided if the Japanese men and the pygmies, or the Japanese men and the Dinkas, were taken together. If mixtures were made up by taking pairs of series at random from all populations in the world, however, the chance of getting a particular pair with means differing by as much as 182 mm. (= Dinka-Japanese mean) is quite small.

This point is illustrated by the interracial distribution given at the bottom of Fig. 2, which was compiled by taking all the means for series of 30 or more men collected by Deniker (1926) and representing populations in all parts of the world. The mean of the distribution is 1646 mm. and the standard deviation 58·8. It can be shown that if pairs of values (i.e. means for series) are drawn at random from it, then the probability that they will differ by less than 80 mm. is 0·56, by less than 120 mm. 0·80, by less than 160 mm. 0·91, and by less than 200 mm. 0·96. It is also known from experience that a mixed distribution of statures made up by combining two samples of equal sizes from populations in which the character is normally distributed, and which differ in their means by 80 mm. is quite likely to give a P on comparison with a normal curve fitted to it which would be high enough to make the incorrect hypothesis thus tested acceptable if

* I am indebted to Dr Otto Samson, who is working on the material, for the statures of the Dinkas. The records were kindly lent by Prof. J. C. Brash.

the source of the material were unknown. With any appreciably wider separation of the means for the two component series all P's found would probably be so low as to indicate clear departure from normality. The position may be summed up roughly in this way in the case of stature: if each of a number of samples of unknown origin was actually a mixture of two samples of equal sizes representing pairs of populations chosen at random from all in the world, then *the form of the distribution alone*, in the case of about half of them, would not be sufficiently peculiar to reveal clearly their distinction from samples representing single populations, but in the case of the other half peculiarity in origin would be indicated. In the same circumstances distributions which were definitely bimodal would be expected to occur occasionally.

Actually, however, such forms are scarcely ever found in practice, and hence it must be inferred that the condition postulated is of an artificial kind. In fact the anthropologist is not at all likely to choose a sample so badly that it represents what may be called a purely "mechanical" mixture of two distinct populations which differ appreciably in their averages for one or more characters. He is far more likely to come across a sample representing a population derived from the partial or complete blood admixture of two groups which were originally distinct. In such cases it is known that the mean for the mixed population will lie between those for the parental groups. Trevor (1938) found that the distributions for groups derived from the crossing of distinct ethnic stocks usually give high P's when compared with normal curves fitted to them. Samples occurring in anthropological practice may relate to diverse kinds of conglomerations, as it were, of mechanical and blood intermixture of populations which would best be considered singly in examining group relationships. In such cases the forms of the distributions of metrical characters will usually fail entirely to indicate the composite nature of the assemblage. A sample of the "white" population of the United States consisting of recruits was found to give a distribution of stature which could be adequately represented by a normal curve (Hoffman, 1918).

Anthropologists find it hard to accept the conclusion that their frequency distributions for heterogeneous material never give any clear indication of the component parts of a particular sample. This situation is observed because the interracial variation of metrical characters is of the same order as, or smaller than, intraracial variation. The situation has been considered in the case of stature, and for this character the interracial distribution gives a standard deviation of the same order as those normally found for intraracial samples. These questions have been investigated most fully for certain measurements of the skull which are more reliable than the vast majority of the measurements available for series of living people. Data similar to those illustrated in Fig. 2 for stature are given in Table III for three cranial characters. The material collected relates to series of adult male skulls made up by 30 or more specimens and

representing populations, believed to be suitable for purposes of classification, which have existed in various parts of the world from about 5000 B.C. to modern times. In the case of each of the three characters considered the distribution for the series with the lowest mean is given first, followed by the distribution of racial means and then that for the series with the highest mean.

TABLE III

Distributions of three cranial measurements, each for an intraracial sample with an extremely low mean (L) and one with an extremely high mean (H), and for an interracial sample (I): male skulls

Cephalic index				Horizontal circumference				Foraminal index			
Central value	*L**	*I*	*H*†	Central value	*L*‡	*I*	*H*§	Central value	*L*‖	*I*	*H*¶
65	7	—	—	450	1	—	—	63·75	1	—	—
67	11	—	—	458	1	—	—	66·25	2	—	—
69	11	3	—	466	3	—	—	68·75	0·5	—	—
71	4	12·5	—	474	8·5	—	—	71·25	0·5	—	—
73	2	13·5	—	482	5·5	—	—	73·75	8	—	—
75	—	33	0·5	490	7	6	1	76·25	7·5	—	—
77	—	26	1·5	498	7	14	1	78·75	5·5	1	1·5
79	—	19	2	506	6	17	5	81·25	2	11	3·5
81	—	12	3	514	3	42	19	83·75	6	51·5	7
83	—	22·5	9	522	—	28	19	86·25	1	19·5	5·5
85	—	7·5	14	530	—	8	34	88·75	1	4	9·5
87	—	1	10	538	—	1	36·5	91·25	0	—	5
89	—	—	7·5	546	—	—	22·5	93·75	2	—	6
91	—	—	6	554	—	—	11	96·25	1	—	4
93	—	—	5	562	—	—	10·5	98·75	—	—	1·5
95	—	—	1·5	570	—	—	4·5	101·25	—	—	1·5
Total	35	150	60		42	116	164		38	87	45
Mean	68·1	77·5	86·4		488·8	512·8	534·6		79·5	84·2	89·2
σ	2·23	4·14	4·11		15·1	10·1	15·7		7·05	1·69	5·63

* Loyalty Islander (Sarasin & Roux, 1916–22).
† Teleghite (Reicher, 1913).
‡ Baining (New Britain) (Bauer, 1915).
§ Reihengräber (Morant, 1928).
‖ Tanganyika (Ried, 1915).
¶ Venezuelian (Marcano, 1890).

The situation is seen to be much the same for the cephalic index as for stature, the interracial variability being rather larger than that for nearly all samples which can be accepted as suitable for purposes of classification. The distributions for the series with extreme means cover contiguous ranges, and there is no overlap. Of all the cranial characters for which adequate data are available the cephalic index is quite outstanding in this respect. For most of them the situation is very similar to that illustrated by the data in Table III for

the horizontal circumference of the skull. In this case the standard deviation for the interracial distribution is appreciably less than any intraracial value found, and the extreme distributions show considerable overlap. The limit in the other direction is exhibited by the index expressing the breadth of the foramen magnum as a percentage of its length. In this case interracial variation is much smaller than intraracial variation, and the interracial distribution falls entirely within the range for any intraracial sample. For most cranial characters the averages for all populations in the world are quite likely to fall within the range provided by any population chosen at random.

It is evident that the chances that a heterogeneous sample will exhibit a distribution diverging appreciably from the normal form are very different for different metrical characters. As far as is known, stature and the cephalic index are the two most likely to show departures from the rule in the case of mixed samples, since the ratio of inter- to intraracial variation is largest for them. For this reason examples relating to the two characters have been given above. Most cranial measurements are decidedly less likely to reveal heterogeneity by the forms of distributions obtained for them, as the difference between the means of any two *samples inadvertently combined compared with the two intra*-group variabilities will be, on the average, appreciably less for them. For some cranial characters interracial variation is so small compared with intraracial variation that no heterogeneous samples will be at all likely to provide peculiar distributions.

In view of the circumstances discussed above, it is not surprising that the distributions which occur in anthropological practice scarcely ever show any appreciable departure from a symmetrical and unimodal form. Detailed statistical analysis of the slight peculiarities of the distributions sometimes observed is never likely to be profitable in the case of small samples—composed of fewer than 300 individuals, say—and in general it does not lead in the case of adequately large samples to any conclusions which aid the classification of the groups represented.

The normal curve may be accepted as the one which almost invariably gives an adequate description of the distributions in the populations sampled, even though some of these populations are unsuitable for the purpose in view. It is usually found to effect this purpose remarkably well, in spite of the fact that the theoretical curve can never be the correct one as it is of unlimited range. In this application it is safest not to draw any inferences regarding the composition of a population from the observed fact that its distributions are normal in form. Very occasionally a clear departure from normality can be taken to indicate that a peculiar and unsuitable population is represented.

For practical purposes the most important conclusion to be derived from the forms of the distributions of anthropometric characters is that dissection of the samples treated into subsamples which might be supposed to represent groups

of different origins is impracticable. Anthropologists in general are unwilling to accept this conclusion. Other characteristics of the samples, such as their variabilities and correlations, also have a bearing on the question and hence discussion of it will be deferred for a later part of this paper.

The writer is indebted to Miss M. L. Tildesley, Prof. E. S. Pearson and Prof. G. von Bonin for criticism which led to the improvement of this first part.

APPENDIX

Certain distributions of cephalic indices provided by Felix von Luschan

As far as the writer is aware, the only published distributions for any anthropometric character which show a clear departure from normality, and which relate to unselected samples made up by more than 100 individuals, are three provided by Felix von Luschan. These are often referred to as evidence that such departures do occur in practice, though usually without comment on the rareness of such an occurrence. All the distributions in question are for the cephalic index, and they all relate to groups of men native to the Near East.

The first is for 179 men who called themselves Greeks measured in Lycia and neighbouring localities in the south of Asia Minor, including a few islands. The figures for the cephalic index were published (von Luschan, 1890). A frequency diagram in the same paper does not agree with these figures, as it shows one index of 95 while the highest in the table is 94. This error was perpetuated when the distribution was reproduced in two later publications (von Luschan, 1911, 1927). The frequencies are shown as histograms in the top diagram of our Fig. 3A.

The second series is made up by 756 Turks from the south of Asia Minor and the north of Syria. The figures for 187 of these men are available (1890) but not those for the remainder. A frequency distribution "reduced to one-third" was given (1911, 1927). It is not possible to read off the individual frequencies from this with absolute accuracy, but the middle diagram of Fig. 3A reproduces them as closely as possible. With reference to this distribution, von Luschan says (1911, p. 236) that it ranges from 69 to 96, but the highest value given in his diagram is 92.

The third series is of 1222 Jews: "52 % of these were Sephardim, whom I measured at Smyrna, at Constantinople, at Makri, and in Rhodes; the rest were Ashkenazim measured by myself...at Vienna, Austria" (1911, p. 226). A frequency distribution "reduced to one-fifth" is given and this is reproduced as accurately as possible in the bottom diagram of Fig. 3A.

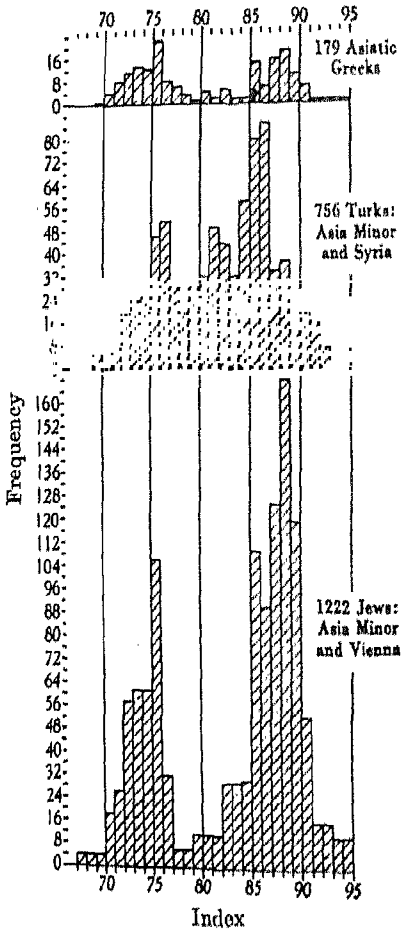
The first and third of these three distributions are clearly bimodal in form. That for the 756 Turks provides less clear evidence of the existence of more than one mode in the population distribution, but it certainly suggests that this must have been clearly asymmetrical. The three distributions provided by von Luschan relate principally to populations of Asia Minor, and their peculiarities might be supposed due to the fact that this region is inhabited by peoples who are racially heterogeneous to an unusual extent. In fact, however, other series from it do not appear to be at all distinguished by any unusual forms of

their metrical characters. Measurements for three series of men which represent the populations discussed by von Luschan as closely as any available are given by:

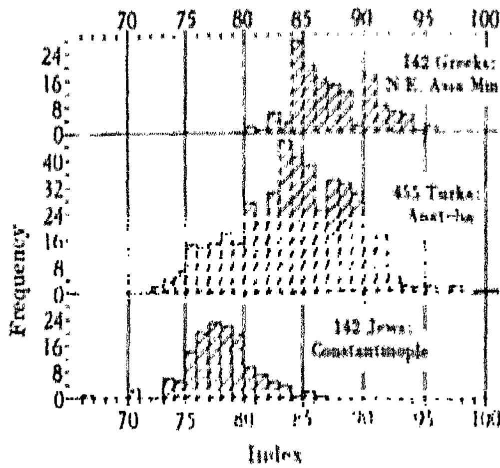
(i) Néophytos (1891): 142 Greeks in the town and neighbourhood of Kerasun in the north-east of Asia Minor;

(ii) Hauschild & Wagenseil (1931): 455 Turks from a number of localities in Anatolia;

(iii) Wagenseil (1925): 142 Jews measured in Constantinople, of whom 107 belonged to that town, 13 came from various countries of Eastern Europe and the remainder were from Asia Minor and Syria.



A. Three distributions provided by F. von Luschan.



B. Three distributions provided by other anthropologists.

Fig. 3. Distributions of the cephalic index for series of men measured in the Near East.

The frequency distributions of the cephalic index for these three series are given in Fig. 3B, and they all differ markedly in form from the corresponding distributions given by von Luschan's data. That for the 142 Greeks is irregular, but little significance can be attached to this fact as the sample is not a large one: it shows no indices below 80, while more than half of von Luschan's Greeks had indices below this value. The measurements for Anatolian Turks and Constantinople Jews do not suggest that the population distributions showed any clear departure from the normal form.

The corresponding pairs of series are also distinguished clearly by their variabilities. The following standard deviations are found:

	Greeks	Turks	Jews
Von Luschan's series	6.96 ± 0.37	5.20 ± 0.13	6.72 ± 0.14
Other series	3.21 ± 0.19	4.76 ± 0.16	2.90 ± 0.17

In comparison with other material, the abnormally high standard deviations of the cephalic index for von Luschan's three series show at once that the samples are of a peculiar kind unsuitable for racial comparisons. This material is altogether exceptional and it does not affect the general conclusion that the distributions of metrical characters for samples treated in anthropological practice almost invariably indicate that the population distributions do not diverge appreciably from the normal form.

REFERENCES

BAUER, L. (1915). "Beiträge zur Kraniologie der Baining (Neu-Pommern)." *Arch. Anthrop., Braunschw.*, N.F. **14**, 145.

DENIKER. J. (1926). *Les Races et les Peuples de la Terre*, 2nd ed. Paris.

DUCKWORTH, W. L. H. (1917). "Notes on some measurements made on subjects in the dissecting-room." *J. Anat., Lond.*, **51**, 167.

ELDERTON, E. M. & WOO, T. L. (1932). "On the normality or want of normality in the frequency distributions of cranial measurements." *Biometrika*, **24**, 45.

FAWCETT, C. D., LEE, A. *et al.*: edited and arranged by PEARSON, K. (1902). "A second study of the variation and correlation of the human skull, with special reference to the Naqada crania." *Biometrika*, **1**, 408.

FIELLER, E. C. (1932). "The distribution of the index in a normal bivariate population." *Biometrika*, **24**, 428.

FÜRST, C. M. & HANSEN, F. C. C. (1915). *Crania Groenlandica.* Copenhagen.

HAUSCHILD, M. W. & WAGENSEIL, F. (1931). "Anthropologische Untersuchungen anatolischen Türken." *Z. Morph. Anthr.* **29**, 193.

HERSKOVITS, M. J. (1930). "The anthropometry of the American negro." *Col. Univ. Cont. Anthrop.* **11.**

HOFFMAN, F. L. (1918). *Army Anthropometry and Medical Rejection Statistics.* Newark, N.J.

HOOKE, B. G. E. (1926). "A third study of the English skull with special reference to the Farringdon Street crania." *Biometrika*, **18**, 1.

HOOTON, E. A. (1925). "The ancient inhabitants of the Canary Islands." *Harv. Afr. Stud.* **7.**

HRDLIČKA, A. (1924). "Catalogue of human crania in the United States National Museum Collections." *Proc. U.S. Nat. Mus.* (12), **63.**

LUSCHAN, F. VON (1890). "Die Tachtadschy und andere Überreste der alten Bevölkerung Lykiens." *Arch. Anthrop., Braunschw.*, **19**, 31.

—— (1911). "The early inhabitants of Western Asia." *J. Roy. Anthr. Inst.* **41**, 221.

—— (1927). *Völker, Rassen, Sprachen.* Berlin.

MACDONELL, W. R. (1904). "A study of the variation and correlation of the human skull, with special reference to English crania." *Biometrika*, **3**, 208.

MARCANO, G. (1890). *Ethnographie Précolombienne du Venezuela. Région des Raudals de l'Orénoque.* Paris.

MATSUMURA, A. (1925). "On the cephalic index and stature of the Japanese and their local differences." *J. Fac. Sci. Tokyo Univ.* Sec. 5, Anthrop. (1), **1.**

MERRILL, A. S. (1928). "Frequency distribution of an index when both the components follow the normal law." *Biometrika*, **20 A**, 53.

MORANT, G. M. (1926). "A first study of the craniology of England and Scotland from Neolithic to early historic times, with special reference to the Anglo-Saxon skulls in London museums." *Biometrika*, **18**, 56.

—— (1928). "A preliminary classification of European races based on cranial measurements." *Biometrika*, **20 B**, 301.

MÜLLER, W. (1905). "Beiträge zur Kraniologie der Neu-Britannier." *Mitt. Mus. Völkerk. Hamburg*, **23**.

NÉOPHYTOS, A. G. (1891). "Le grec du nord-est de l'Asie Mineure au point de vue anthropologique." *Anthropologie, Paris*, **2**, 25.

PEARSON, KARL (1897). *The Chances of Death and Other Studies in Evolution*. London.

PEARSON, K. & DAVIN, A. G. (1924). "On the biometric constants of the human skull." *Biometrika*, **16**, 328.

PITTARD, E. (1909–10). "Crania Helvetica. 1. Les crânes valaisans de la vallée du Rhône." *Mém. Inst. nat. genev.* **20**.

QUETELET, A. (1835). *Sur l'Homme et le Développement de ses Facultés ou Essai de Physique Sociale*. Paris.

—— (1871). *Anthropométrie ou Mesure des Différentes Facultés de l'Homme*. Brussels.

REICHER, M. (1913). "Untersuchungen über die Schädelform der alpenländischen und mongolischen Brachycephalen." *Z. Morph. Anthr.* **15**, 421.

RIED, H. A. (1915). "Zur Anthropologie des abflusslosen Rumpfschollenlandes im nordöstlichen Deutsch-Ostafrika." *Abh. Hamburg. Kolon. Inst.* **31**.

SARASIN, F. & ROUX, J. (1916–22). *Nova Caledonia: Forschungen in Neu-Caledonien und auf den Loyalty-Inseln*. C. *Anthropologie*. Berlin.

SCHIFF, F. (1912). "Beiträge zur Kraniologie der Czechen." *Arch. Anthrop., Braunschw.*, N.F. **11**, 253.

SCHMIDT, E. (1886). *Die anthropologischen Sammlungen Deutschlands*. Leipzig Cat.

ŠEBESTA, P. & LEBZELTER, V. (1933). "Anthropology of the central African pygmies in the Belgian Congo." *Czech. Acad. Sci. Arts*, 2. Class, *Anthropologica*.

STIEDA, L. (1883). "Über die Anwendung der Wahrscheinlichkeitsrechnung in der anthropologischen Statistik." *Arch. Anthrop., Braunschw.*, **14**, 167.

TOCHER, J. F. (1924). "Anthropometric observations on samples of the civil populations of Aberdeenshire, Banffshire, and Kincardineshire", and "A study of the chief physical characters of soldiers of Scottish nationality, etc." *The William Ramsay Henderson Trust Reports*, 2 and 3.

TOPINARD, P. (1885). *Eléments d'Anthropologie Générale*. Paris.

TREVOR, J. (1938). "Some anthropological characteristics of hybrid populations." *Eugen. Rev.* **30**, 21.

WAGENSEIL, F. (1925). "Beiträge zur physischen Anthropologie der spaniolischen Juden und zur jüdischen Rassenfrage." *Z. Morph. Anthr.* **23**, 1925.

WITT, H. (1879). "Die Schädelform der Esten." *Diss.* Dorpat.

A STUDY OF THE CRANIAL AND OTHER HUMAN REMAINS FROM PALESTINE EXCAVATED AT TELL DUWEIR (LACHISH) BY THE WELLCOME-MARSTON ARCHAEOLOGICAL RESEARCH EXPEDITION

By D. L. RISDON
Crewdson Benington Student

CONTENTS

		PAGE
1.	Introduction	99
2.	The general nature of the remains and remarks on the origin of the ossuaries	102
3.	Remarks on the condition and anomalies (other than dental) of the Lachish crania	106
4.	Remarks on the condition and anomalies of the jaws of the Lachish skulls	117
5.	The statistical nature of the material	121
6.	The relationships of the Lachish and Ancient Egyptian populations judged from comparisons of mean measurements	128
7.	Comparisons of the Lachish and Ancient Egyptian series for characters considered singly	143
8.	The contours of the Lachish series	145
9.	Comparisons of the Lachish series with series of crania and living people other than Egyptian	158
10.	Mandibles and long bones of the Lachish series	159
11.	Summary and conclusions	160
	Appendices I. Definitions of skull measurements taken	162
	II. Tables of individual measurements	*to face p.* 162
	List of Plates	163
	References	165
	Plates I–XXIII	*to face p.* 166

Note that tissues of Figs. 4–9 are contained in end pocket

1. Introduction

The human skeletal remains which form the subject of this report were excavated by the Wellcome (later the Wellcome-Marston) Archaeological Research Expedition to the Near East from 1933 to 1936. They were collected by the late Mr J. L. Starkey, who was then the Director of the Expedition, and it was owing to his enthusiasm and untiring efforts in the field that so much material is now available for study. His tragic death in 1938 deprived archaeology and physical anthropology alike of one who was chiefly responsible for the collection of some of the most valuable evidence extant relating to the early history of the peoples of the eastern Mediterranean.

The human remains from Lachish were brought to England, and the greater part of the work on them was carried out at the Galton Laboratory, University College, London, under the supervision of Dr G. M. Morant. I am indebted to the Trustees of the late Sir Henry Wellcome for a grant which enabled me to carry out this work over a period of rather more than three years, and also for a subsidy which has made possible the complete presentation of the illustrative and tabular matter collected for the report. I also owe a very great debt of gratitude to Dr Morant for his untiring help and advice in all the stages of its preparation. A complete report of the excavations is in course of preparation, and I have to thank Mr Charles Inge, the present Director of the Expedition, for most of the following particulars relating to the discovery of the bones.

The only published account of them is a note by Mr Starkey (1936), prefaced to a description of three trepanned skulls by Dr Wilson Parry. He reports that a roughly circular chamber (No. 107), which contained a deposit of human remains much damaged by fire, was opened in 1934. At the same time an adjoining and larger rectangular cavern (No. 120) was located. Starkey writes:

"The top layer consisted of many animal bones, mostly pig, and this refuse should be ascribed to the latter half of the Judean kingdom....The lower or main deposit consisted of a mass of human bones, the remains of at least 1500 bodies. As they were pitched in through the hole in the broken roof, the skulls rolled down from the apex of the pile to the sides of the chamber."

Sherds of pottery intermixed with the bones can all be assigned to the seventh and eighth centuries B.C.

"Some bones were partially calcined, suggesting that they were abstracted from burnt buildings....Careful supervision of the clearance failed to establish that any crania were in articulation with vertebrae, and the jaws were rarely attached: in fact, no order was seen in the jumbled mass."

It is suggested that the ossuary was probably connected with the salvage of Lachish after its partial destruction by Sennacherib, King of Assyria, in 701 B.C.

"When floor level was reached it became clear that the tomb had been previously used as a dwelling, a door had been cut at the N.E. [given erroneously as N.W. in the published account] corner, connecting the smaller circular tomb with it, which contained the 500 bodies discovered in 1934. From the style of both chambers it is certain that they were originally excavated to contain early fifteenth-century burials."

Excavation of the area was extended and two other tombs were found, of which no details have yet been published. Fig. 1 is a plan of the area in question, showing all four tombs from which the human remains treated in this study were obtained.

The archaeologists report that:

"All the tomb chambers were adjoining and interconnected, and the deposits in Nos. 107, 108, and 120 were identical and as described by Mr Starkey.

Tomb No. 116 is in a slightly different category; it is an artificial cavern, with a small entrance cut in the south wall of No. 108 near the bottom, containing normal burials comparatively undisturbed. The period however is the same."

It is said that the later discoveries do not provide any evidence which makes closer dating possible. They include a Cypriote juglet, of a type common in Palestine in the eighth century B.C., scarabs and scaraboids, bone pendants, and blue glaze amulets. The last include representations of Egyptian deities

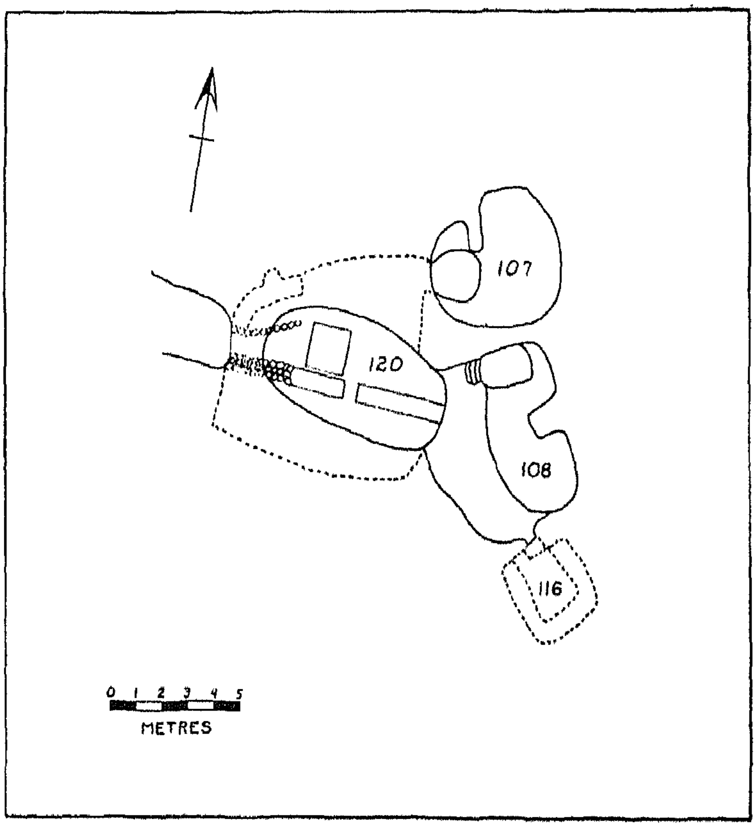


Fig. 1. Plan of the tombs at Lachish from which the human remains were recovered. The broken lines indicate approximately the floor area of Tomb 120, and the outline of the benches (inner line) and the maximum perimeter of Tomb 116. The approximate heights of the tombs were: 107, 2 m.; 108, 2·2 m.; 116, 1·6 m.; and 120, 3·6 m.

which were long popular in Palestine. There is nothing in the objects that one would be surprised to find in normal burials at Lachish in the middle of the Jewish period.

Further particulars regarding the state in which the human remains were found are available. In all the tombs, except No. 116, there were distinct layers of animal bones above the human deposits, and in the shallower part of Tomb 108 they came up to the roof level, the roof itself having been broken away. In Tomb 107 the apex of the pile of human bones was about 1 m. above the floor

level, and the same height was about 1·3 m. in Tomb 120. There were fewer human remains in Tomb 108: these were more scattered, and there were many animal bones above them.

Tomb 116 is in a different category from the others, for it contained several formal burials laid out on the benches, with the usual funerary equipment. These burials were partially disturbed, and the skulls could not be distinguished from others which may have rolled in at the time of the filling of Tomb 108. The roof of Tomb 116 was intact, and most of the long bones collected come from it as they were the best preserved.

The estimate that the remains of 1500 individuals were represented applies to Tomb 120, and it is admittedly very rough. It was impossible to judge whether the piles of bones included the complete skeletal remains of 1500 people or not, but the impression was gained that the skeleton of the head was far better represented than that of any other part of the body. A few mandibles attached to crania were found, but there appeared to be no other cases of parts of the skeleton existing in proper articulation. A suggested explanation of the circumstances which led to the existence of the ossuary is given in the following section of this paper, as evidence derived from the bones preserved has a bearing on this topic.

Plates I and II reproduce photographs taken in the interior of Tomb 120 after the clearance down to floor level. The piles of skulls which had accumulated round the walls are shown *in situ*. Plate II B is a close view of a small group of skulls seen to the left of the beam in Plate I A and above the pick in Plate I B.

Other human remains from a number of tombs at Lachish, representing fewer than 100 individuals in all, have been preserved, and the writer is preparing a separate report on this material.

2. The general nature of the remains and remarks on the origin of the ossuaries

The human remains described in this report comprise all that were preserved from the four tombs referred to in the foregoing section. They are made up principally of crania, which appeared to be more abundant in the deposits than any other parts of the skeleton, but there are also series of mandibles and long bones of the limbs, and a few other bones. A selection was made of the better preserved remains, and preference was given to crania. It should be realized that the relative numbers of different parts in the collection preserved cannot be supposed to be proportional to the frequencies with which they actually occurred in the ossuaries.

The specimens were dipped in paraffin wax in the field, and the cleaning of them in the laboratory was a laborious task. While it was being carried out, four crania distorted by earth pressure were found, as well as a certain number of fragments of other parts of the skeleton which may have belonged to individuals

included in the series, or to individuals not otherwise represented. All this material was discarded, and it is not counted in the totals. The numbers of specimens, most of which are more or less incomplete, from each tomb are given in the table below, no distinctions of sex or age being made:

	Tomb				
	107	108	116	120	Totals
Crania	74	9	45	567	695
Mandibles	12	1	7	56	76
Femora	2	—	50	28	80
Tibiae	2	1	27	15	45
Fibulae	—	—	1	—	1
Humeri	2	—	23	23	48
Radii	1	—	5	8	14
Ulnae	3	3	3	6	15
Sacra	—	—	—	2	2
Clavicles	4	—	2	1	7

There are also a few vertebrae associated with skulls.

The absence of other bones of the skeleton which would normally be well preserved, such as those of the wrist and ankle, should be noted. In this connexion Mr Inge remarks: "Less attention was paid to the skulls of children owing to their supposed smaller anthropometrical value, and the same applies to parts of the skeleton other than skulls and long bones."

None of the bones are known to be associated, except that ten mandibles and crania were preserved together, and are undoubtedly paired. It is probable that a considerable proportion of the other mandibles do belong to the crania, but it is not possible to associate them. Experience has shown that any attempt to pair parallel series of these parts is unprofitable. It was also impossible to associate the other bones with the crania or with one another.

After separating the immature specimens, the crania and mandibles were sexed anatomically, giving the totals shown in the table below. Estimates of

		Tomb				Totals
		107	108	116	120	
Crania	Adult ♂	33	5	19	303	360 (51·8%)
	Adult ♀	31	3	25	215	274 (39·4%)
	Immature	10	1	1	49	61 (8·8%)
Mandibles	Adult ♂	5	0	3	26	34 (44·7%)
	Adult ♀	7	1	4	24	36 (47·4%)
	Immature	0	0	0	6	6 (7·9%)

this kind cannot be exact in all cases, of course, but experience suggests that 80–90 % are likely to be correct. The inclusion of a certain proportion of male-like, but actually female, specimens in the male series, or vice versa, is not likely to affect statistical constants appreciably. It has been shown by Martin (1936) and Cleaver (1937), that isolated mandibles can be sexed anatomically as accurately as crania. Particulars regarding the other bones of the skeleton are given in § 10 below.

The proportions of the sexes and of immature individuals are seen to be very much the same in all the tombs. The corresponding male and female percentages in the table are not particularly close, but the series of mandibles is obviously too small to give reliable proportions. It is safe to assume that there are rather more males than females in the collection preserved. This does not necessarily imply that the same was true for the total number of individuals deposited in the tombs, as the male skulls, being stronger on the average, would stand a slightly better chance of being preserved than female specimens. It is probable, for a similar reason, that the proportion of immature individuals was very considerably higher in the tombs. The proportions of men, women and children there appear to have been very similar to those which would be expected in a normal cemetery population.

Estimates of the ages at death of the adults are discussed in § 3 below. It is shown there that the population was considerably younger than such as is normally found in ancient or recent cemeteries. There are very few aged specimens in the collection.

The condition of the bones throws little further light on the question of how and why they were deposited in the tombs, but the following observations may be noted in this connexion. There is only one skull (No. 108, ♂, Tomb 120) with wounds which are likely to have been the cause of death. This has pieces of bone removed from both parietals (see Plate III B), and a long cut extending across the left temporal line. The other lines on the left parietal and extending across the sagittal suture, along which the bones are broken, may have been made at any time after death. As far as can be told the damage shown in the case of all the other specimens occurred in the tombs, and after the deaths of the individuals. The greater part of it may be attributed to the crushing weight of the piles of bones.

Among the totals for all tombs combined, there are twenty-nine (8·1 %) adult male, eight (2·9 %) adult female, and two (3·3 %) immature skulls with wounds, mostly slight, which had healed completely during life. These frequencies are not higher than those commonly found for excavated series of crania.

In his account of the discovery of the remains, Starkey notes that those in Tomb 107 were much damaged by fire, and that some bones in Tomb 120 were "...partially calcined, suggesting that they were abstracted from burnt

buildings". It is probable that the great majority of the burnt specimens were too damaged to be worth preserving. In the series preserved there are only three skulls with burnt patches.* No. 485 (♀, Tomb 107) has an extensive area on the occipital and right parietal bones affected (see Plate III A), No. 537 (♀, Tomb 120) shows a smaller area on the occipital bone, and No. 681 (immature, Tomb 120) shows a small burnt patch above the right orbit.

It may be noted, too, that there are three trepanned skulls (all males from Tomb 120), and a few artificially deformed skulls from the same tomb. There are two male skulls and one female skull which show definite signs of diseased conditions affecting the bone. Again, they all come from Tomb 120 but it must be remembered that the series from this tomb is far longer than that from any of the other tombs.

The evidence relevant to the provenance of the human remains from the four tombs at Lachish consists partly of archaeological observations, and partly of deductions from the remains preserved. It must be remembered that the series available were selected from a larger bulk of material. Taking the four tombs together, the total remains in hand must represent rather more than 700 individuals. It was estimated that those of about 1500 were present in Tomb 120 when it was opened. As the dimensions of the piles of bones are known approximately, it may be asked whether they could possibly have been made up by bodies, or skeletons, which were complete when deposited. Taking Tomb 120 alone, the dimensions of the pile were roughly: height $1{\cdot}3 \times 7{\cdot}5 \times 8{\cdot}5$ m. (see Fig. 1), giving a volume of about 83 cu.m. It is possible that at least 1500 *complete* skeletons, including some of children, were interred in a space of this size. The tomb was nearly three times as high as the pile of bones, and hence it is also possible that it could have contained a heap of 1500 complete bodies, which might be supposed to have subsided on disintegration. The animal bones, which formed a distinct layer over the human remains, may have been thrown in after such a subsidence. So far there is nothing to preclude the possibility of the interment having been either of bodies, or of complete skeletons, such as might have occurred in clearing a neighbouring cemetery.

The hypothesis that the remains were obtained from a cemetery—so that Tomb 120, and possibly Nos. 107 and 108 as well, were really clearance pits—seems to be unacceptable for several reasons. In the first place, it is almost invariably found that among bones cleared from cemeteries, crania and femora are far better represented than any other parts of the skeleton, and in such a case a pile as large as that in Tomb 120, say, would have been likely to contain far more than 1500 skulls. After clearance, too, it would have been unlikely that any skulls and mandibles would have been associated, though a few pairs were

* It was found that the paraffin wax in which the bones had been dipped could be removed most effectively with the aid of a flame, and although great care was exercised, a few specimens now show signs of slight burning, chiefly at the base of the skull, acquired in the laboratory.

found. Another argument against cemetery clearance is that few skeletons of children are likely to have been thrown in the tombs, on the hypothesis considered, and practically all these would probably have been too crushed in the pile to be worth preserving. Of the total number of skulls in the series, 8·8 % are immature, including a fair proportion of children under 10 years of age, and it is unlikely that this frequency would have been so high after two stringent selections of the material. The scarcity of aged specimens also tells against the hypothesis of clearance from a cemetery.

In view of the fact that the skeletons of children are far more easily damaged than those of adults, and that bones of women are more easily crushed than those of men, it is not at all unlikely that the proportions of immature to mature, and of female to male, skeletons in the tombs, were very similar to those of the *living* population of Lachish at a particular time. The rarity of aged specimens must be remembered. These facts seem to tell against the theory that the individuals in the tombs were massacred, but a stronger argument against it is the fact that only one skull showing injury which might have been the cause of death is recorded.

The most plausible explanation, which appears to be in accordance with all the evidence, is the following. It may be supposed that some catastrophe, such as pestilence or earthquake, overtook the population of Lachish about the year 700 B.C., and that a large proportion of the inhabitants were victims. Ordinary burial at such a crisis would have been impossible, and in clearing the town, some time after it, the underground chambers in question would have been convenient depositories into which bodies could have been thrown. The fact that some were burnt is not surprising, as the catastrophe may have been accompanied, or followed, by a fire in part of the town. The fact that the age distribution is appreciably younger than would be anticipated for any cemetery population of the period is accounted for on the hypothesis considered.

3. Remarks on the condition and anomalies (other than dental) of the Lachish crania

Remarks on the Lachish skulls are given in the appended tables of individual measurements. Owing to the incomplete nature of many of the specimens, the totals observable in the case of particular anomalies are generally considerably less than the totals for the whole collection, and hence they have to be given separately in considering the frequencies of different conditions. The total numbers of the skulls from different tombs divided into sexes for the adults, and the numbers of immature specimens, are given on p. 103 above. In general, the frequencies are considered separately in the case of the long series from Tomb 120, and of the series made up by combining the skulls from all the other tombs. Subdivision of the latter is inadvisable, as the numbers from all the tombs other than 120 are small.

Sutures. Remarks on the coronal, sagittal, and lambdoid sutures are given in the tables of individual measurements. If no mention of one of these is made there, this indicates that it shows no sign of closing. Every specimen is complete enough to observe the three principal sutures. The approximate estimate of the age constitution of the series, which can be obtained from these observations, will be considered first. Frequencies are given in Table I. If a suture showed any sign of the beginning of closure, the skull was counted in the second category (sutures beginning to close or partly closed). A suture was counted as closed if synostosis was apparent for at least the greater part of its length.

TABLE I

*The age constitution of the adult Lachish and other series estimated from the state of closure of the principal calvarial sutures (coronal, sagittal, and lambdoid)**

	Series	All sutures open	Sutures beginning to close or partly closed	All sutures closed	Totals
♂	Lachish, Tomb 120	123 (43·3%)	153 (53·9%)	8 (2·8%)	284
	Other tombs†	26 (45·6%)	28 (49·1%)	3 (5·2%)	57
	Total‡	149 (43·7%)	181 (53·1%)	11 (3·2%)	341
	Kerma, 12th and 13th dyn.	16 (11·3%)	83 (58·9%)	42 (29·8%)	141
	Gizeh, 26th–30th dyn.	74 (37·0%)	103 (51·5%)	23 (11·5%)	200
	Farringdon St, Londoners	19 (12·2%)	100 (64·1%)	37 (23·7%)	156
	Whitechapel, Londoners	21 (15·4%)	85 (62·5%)	30 (22·1%)	136
	Spitalfields, Londoners	103 (19·3%)	301 (56·3%)	131 (24·5%)	535
	Hythe	23 (20·9%)	62 (56·4%)	25 (22·7%)	110
♀	Lachish, Tomb 120	154 (74·0%)	54 (26·0%)	0 —	208
	Other tombs	45 (76·3%)	14 (23·7%)	0 —	59
	Total	199 (74·5%)	68 (25·5%)	0 —	267
	Kerma, 12th and 13th dyn.	40 (35·1%)	51 (44·7%)	23 (20·2%)	114
	Farringdon St, Londoners	94 (46·3%)	80 (39·4%)	29 (14·3%)	203
	Whitechapel, Londoners	66 (46·8%)	49 (34·8%)	26 (18·4%)	141
	Spitalfields, Londoners	121 (49·4%)	94 (38·4%)	30 (12·2%)	245
	Hythe	51 (59·3%)	26 (30·2%)	9 (10·5%)	86

* The data in this table for all the series other than the Lachish are taken from Stoessiger & Morant (1932, p. 170) and Collett (1933, p. 259).

† Comprising Tombs 107, 108, 116.

‡ The posthumously and artificially deformed Lachish skulls, and those not included in the pooled series on account of the fact that they show premature closing of the sagittal suture, as well as No. 380 (premature closing of the coronal suture) and No. 382 (of unusual form), are not included in the totals given in this table.

It should be noted that all the data relating to sutures refer to their ecto-cranial aspects. The endocranial surface of the brain-box could only be observed

in a few specimens, as cleaning was impossible, and hence no remarks on the inside of the skull were recorded.

The division of the adult skulls into the three categories shown in Table I was made with the primary purpose of obtaining a rough estimate of the distribution of age of death for the individuals whose skulls are preserved. It is known that there is considerable variation in the ages at which the different sutures close; a skull with the sutures partly closed, for example, may well have belonged to an older individual than one with all sutures open. The actual age distributions for the separate groups, if they could be known, would doubtless show very considerable overlap. In spite of this it is safe to assume that the percentages may fairly be compared with those given for another series observed in precisely the same way, in order to determine differences in the age constitutions of the samples. The data for the other series given in the table were actually collected by Dr Morant, who confirmed the fact that my records had been obtained in a comparable way.

There are very clear differences between the percentages for the Lachish skulls and those for the other series, while the corresponding values for Tomb 120 and the combined series from the other tombs are remarkably close. The usual

TABLE II

*The frequencies of different orders of closing of the principal calvarial sutures for the Lachish skulls (all tombs)**

	♂	♀
Sagittal closing first, coronal and lambdoid open or closing together	74†	23‡
Sagittal closing first, coronal second and lambdoid last	29	3§
Sagittal closing first, lambdoid second and coronal last	17‖	2
Coronal closing first, sagittal and lambdoid open or closing together	21¶	24
Coronal closing first, sagittal second and lambdoid last	3	6
Lambdoid closing first, sagittal and coronal open or closing together	2	1
Lambdoid closing first, sagittal second and coronal last	2	0
Sagittal and coronal closing together before lambdoid	37	12
Sagittal and lambdoid closing together before coronal	11	1
All three sutures closing together	7	2
Totals	203	74

* Excluding the posthumously and artificially deformed, but including the eleven male and six female skulls showing premature closing of the sagittal suture, and No. 380, showing premature closing of the coronal suture.

† Including eight showing premature closing of the sagittal suture.

‡ Including five showing premature closing of the sagittal suture.

§ Including one showing premature closing of the sagittal suture.

‖ Including three showing premature closing of the sagittal suture.

¶ Including No. 380.

sexual difference, due to the fact that the sutures close at a later age for females than for males, is found. It is evident that the average age at death must have been considerably less for the Lachish adults than for any other of the groups of people. All the other series, except the Whitechapel (obtained from a burial pit, probably used in time of plague), are believed to represent ordinary cemetery populations.

Statistics relating to the orders in which the three principal calvarial sutures were closing are given in Table II. It was possible to observe these orders in the case of all the adult skulls for which closure had commenced, except one aged male (No. 357) in which all three sutures are completely obliterated. The special series of adult specimens presumed to have been affected by premature closing of the sagittal suture, and the one (No. 380) distorted owing to premature closing of the coronal suture, are included in Table II. They were omitted from Table I, as their anomalous conditions might give fallacious estimates of age at death.

The figures show that for the majority of skulls the sagittal suture began to close before the other two, and that the coronal normally closed before the lambdoid. The order sagittal-coronal-lambdoid is also found with the greatest frequency for the Kerma Egyptian series (Collett, 1933). The frequencies for the two series may be summarized in another way, as follows, only skulls for which the first suture to close can be observed being included:

	♂		♀	
	Lachish	Kerma	Lachish	Kerma
Sagittal closing before other two	120 (81·1%)	75 (79·8%)	28 (47·5%)	34 (68·0%)
Coronal closing before other two	24 (16·2%)	18 (19·1%)	30 (50·8%)	14 (28·0%)
Lambdoid closing before other two	4 (2·7%)	1 (1·1%)	1 (1·7%)	2 (4·0%)
Totals	148	94	59	50

The values of the corresponding percentages for the male series are quite insignificant, but differences which must be considered significant are found for the first two pairs of percentages in the case of the females. It should be pointed out that the question which suture shows a more advanced stage of closure than the other two may depend on the age of the individual, and if the rates at which the different sutures close are markedly different comparisons between the frequencies considered for different series may be misleading.

For the English series, examined in the same way, the sagittal suture was found to be the first to close, almost invariably followed by the coronal and lambdoid, which appeared to close together. For a Negro series the coronal

was found to show a slight tendency to close before the sagittal. There are certainly racial differences with regard to the order in which the calvarial sutures close, and, as far as can be seen, the Lachish series agrees with Ancient Egyptian in this respect.

It can be seen from Table II that there are only four male and one female skull showing the lambdoid suture closing before either of the others. An examination of these specimens shows that their sutures were definitely closing in an anomalous way in the majority of cases. Of the four males, Nos. 184, 265 and 330 show the occipito-mastoid suture obliterated on both sides; No. 184 also shows the parieto-mastoid suture obliterated on the left. No. 265 shows the temporal squama completely fused to the parietal on the right, and partly fused on the left; and No. 330 shows both parieto-mastoid sutures obliterated and the temporal squamae largely fused to the parietals. The other male skull with the lambdoid suture closing first (No. 268) has all the sutures between the temporal bones on the one hand, and the occipital and parietal, on the other, open. The female specimen (No. 619) has the occipito-mastoid sutures obliterated, and the posterior parts of the temporal squamae fused to the parietals.

There are five other male skulls, and two other female, showing obliteration of the parieto-mastoid suture, and/or partial or complete fusion of the temporal squamae to the parietal bones. In these cases the principal calvarial sutures are open or closing in a normal order, as far as can be seen, and one (No. 357) have them obliterated. The specimens in question are:

No. 90, ♂. Occipito-mastoid suture *L* obliterated, and posterior half of temporal squama *L* fused to parietal.

No. 171, ♂. Occipito-mastoid suture *L* obliterated, and posterior half of temporal squama *L* fused to parietal.

No. 230, ♂. Temporal squama *R* largely fused to parietal.

No. 262, ♂. Occipito-mastoid suture *L* and parieto-mastoid suture *L* obliterated.

No. 357, ♂. Temporals completely fused to occipital and parietal bones.

No. 383, ♀. Occipito-mastoid sutures, *R* and *L*, and parieto-mastoid suture *L* obliterated.

No. 478, ♀. Occipito-mastoid sutures closed, *R* and *L*, temporal squamae fused to parietals, *R* and *L*.

There is one juvenile specimen (No. 680) showing the right parieto-mastoid suture obliterated, but all others open.

In all there are six male skulls showing the temporal squamae completely, or partially, fused to the parietal bones, three on both sides, three on the left only, and one on the right only; and there are two female skulls showing the

same condition on both sides. A count was made of the number of cases showing the occipito-mastoid suture completely obliterated on one or both sides; the totals are:

	Right and left	Left only	Right only	Total no. of adult skulls examined
♂	6	8	1	351
♀	6	2	1	272

When the condition is unilateral, it appears to be shown more frequently on the left than on the right side, but the numbers are too small to warrant a generalization.

Premature closing of the sagittal suture was noted in the case of eleven male and six female adult skulls, all coming from Tomb 120. These were not included in the series used for comparative purposes, but individual measurements for them are given in a separate section of the appended tables. In the case of six male and four female specimens the sagittal suture is completely obliterated, and the coronal and lambdoid completely open. In the case of the remaining five male and two female specimens the sagittal is completely obliterated and one or both of the other two sutures are closing or closed. Few of these skulls show any clear sign of distortion owing to the abnormal closure of the sagittal suture, the only exceptions being No. 364 (male) and Nos. 667 and 672 (female), these having the lowest cephalic indices. The three can safely be called scaphocephalic, but the forms of the others are not exceptional, and hence it may be presumed that their sagittal sutures did not become synostosed until growth of the brain-box was nearly completed. The cephalic indices of all the specimens showing premature obliteration of the sagittal suture compared with those of the normal series (Tombs 107, 108, 116 and 120 combined) are as under:

		60·05–65·05	65·05–70·05	70·05–75·05	75·05–80·05	80·05–85·05	85·05–90·05	Totals
♂	Normal series	1	26	160	112	11	0	310
	Premature closing of sagittal suture	1*	1	7	2	0	0	11
♀	Normal series	0	7	112	117	15	1	252
	Premature closing of sagittal suture	1*	1*	3	1	0	0	6

* Accepted as scaphocephalic.

There are no brachycephalic skulls among those in the special series, but the indices for the majority of them are clearly not abnormal. The fact that all the

adult specimens showing premature closing of the sagittal suture came from the same tomb (No. 120) is suggestive. It may well be that the condition is of genetical origin, and that the affected individuals were related. This explanation would account for the high proportion of specimens exhibiting the anomaly.

One other female skull (No. 577) has the posterior half of the sagittal suture completely obliterated, but the anterior half, and the coronal and lambdoid sutures, completely open. The specimen also shows a distinct post-coronal depression, though this is not apparent in the median sagittal plane.

There is only one other skull (not included in the normal series) which apparently shows deformation resulting from premature closing of a suture. This is No. 380 (Plate VII F): its coronal suture is almost obliterated, while the sagittal and lambdoid are completely open. This is probably a case of oxycephaly, and, though the deformation is not unlike that which was clearly produced artificially in the case of certain specimens, yet all those showing artificial deformation most clearly have the coronal suture completely open. No. 380 has a height-length ($100\ H'/L$) index of 83·3, whereas the highest for a normal adult male skull of the series is 80·8.

No. 59 (male) shows the superior half of the coronal suture on the right side obliterated, but other parts of this suture open, while the sagittal is beginning to close and the lambdoid is open. It does not appear to be deformed, the height-length index of 75·1 being rather high but by no means extreme. No. 584 (female) shows the right side of the coronal suture closed, while the left side and the sagittal and lambdoid sutures are open.

The complete metopic suture was found for twenty-six of the 341 male adult skulls included in Table I, and slight traces of it were observed in a few other cases: it was also found for twenty-two of the corresponding total of 267 female specimens. There are nineteen of the male, and twenty-one of the female affected individuals showing the metopic and sagittal sutures either both open, or closing together. In the remaining female, and six of the remaining male specimens, either one or the other suture is closing first, but the difference between the state of closure of the two is only slight. The greatest difference in this respect is shown by the last male specimen, which has the sagittal suture closed and the metopic open. The conclusion derived from earlier material that the metopic suture normally closes about the same time as the sagittal, if it persists to an adult stage, is thus confirmed.

The male percentage for the condition is 7·6, and the female 8·2. These values are higher than those for the Kerma Egyptian (4·5 male, and 6·2 female), but rather lower than the percentages usually found for European series. Of the sixty-one immature skulls from Lachish, there are eleven (18·0 %) showing the complete metopic suture, an appreciably higher proportion. Of the eleven male adult skulls not included in the normal series on account of premature closing of the sagittal suture, one is metopic, and of the six female in the same group

two show the condition. Of the seven male skulls which show patent artificial deformation, or a suspicion of this condition, two are metopic, and of the two female, one is metopic. The two specimens which are most clearly deformed (Nos. 381 male, and 673 female) are both metopic, which is curious.

A few other sutural anomalies, besides those treated under *supernumerary bones* below, were noted. Among 267 female skulls, twelve were found with traces of the suture between the ex- and supra-occipital bones, but no trace was found in any male specimen. Only one example of a complete suture across a malar bone was found (No. 154 male), the right malar bone being divided and the left normal. No. 216 (male) shows incomplete fusion of the basi-occipital and basi-sphenoid, although it is almost certainly adult. No. 491 (female) shows the suture between the frontal and sphenoid bones obliterated on both sides, although the coronal suture is open. Several examples of fused nasal bones are noted in the remarks.

The region of the pterion and supernumerary bones. The region of the pterion was examined on all the skulls, and the totals for which the sutures there are visible are given in the table below. Epipteric bones are recorded in the remarks on individual specimens, and, as usual, they are found to be diversified in number and size. The numbers of specimens showing contact between the temporal and frontal bones are given in the following table, the right and left sides being considered separately:

	♂		♀		Immature	
	R	*L*	*R*	*L*	*R*	*L*
Total no. of skulls with sutures at pterion visible on the side in question*	257	258	136	138	61	61
No. of skulls with fronto-temporal articulation	2	3	3	3	1	2
No. of skulls with pterion in *K*	—	—	1	2	—	—

* For the normal series (all tombs), excluding the posthumously and artificially deformed skulls, those showing premature closing of the sagittal or coronal suture, and No. 382. Among all these specimens excluded there is one case of contact between the temporal and frontal bones.

Of the four male skulls with fronto-temporal articulation, one shows the condition on the right side only, two show it on the left side only, and the last shows it on both sides. Of the five female skulls, two show the condition on the right side only, two on the left side only, and for the last it is bilateral. Of the two juvenile specimens, one has fronto-temporal articulation on the left side only, and for the other the condition is bilateral. Of the two female specimens with the pterion in *K*, one case is on the left side only, and for the other the condition is bilateral. The low frequency of cases showing contact between the temporal and frontal bones is comparable with those recorded for European

series. One of the male skulls suspected to be artificially deformed shows fronto-temporal articulation on both sides, but all the others have normal pteria.

Examples of one or more wormian bones between the temporal squamae and parietals were observed with a frequency which appears to be unusually high. There are ten cases of the condition exhibited unilaterally, and five bilaterally, among the male skulls; four unilaterally and four bilaterally for the female; and one unilaterally and one bilaterally among the immature specimens.

The normal series of 341 male skulls (all tombs) includes the following cases of complete or partial division of the occipital bone: *os épactal* (preinterparietal) 2, interparietal bones 8 (only *os pentagonale* and *os triangulare R* separate 2; only *os pent.* separate 2; only, *os tri. R* separate 2; only *os tri. L* separate 1; two large bones meeting below lambda 1—No. 60, Plate XIV B), traces of horizontal suture of interparietal bones near asteria 11. The normal series of 267 female skulls includes the following cases: *os épactal* 3, interparietal bones 2 (only *os tri.* separate 1; only *os pent.* separate 1), traces of horizontal suture of interparietal bones 9. The series of 61 immature specimens include the following cases: *os épactal* 1, only *os pent.* separate 1. All the skulls not included in the normal series, on account of deformation or for other reasons, have normal occipital bones. The percentages of the occurrence of true interparietal bones, of one form or another, are 2·3 for the male adults, and 0·7 for the female adults. These values are close to those given for the Kerma Egyptian series (Collett, 1933, p. 265), viz. 3·6 and 0·9 %, respectively.

One male and one female skull with an ossicle of bregma were noted in the total series. Ossicles of lambda of varying sizes were far more frequent. In general the sutures appeared to be moderately complicated and wormian bones in the lambdoid suture are by no means rare. A few cases of supernumerary bones in the sagittal and coronal sutures were noted, but these are all small, with one exception. This is a male specimen (No. 299), the sagittal suture being closed and the coronal closing. There is a large supernumerary bone in the right side of the coronal suture above the pterion. It is roughly triangular in shape (see Plate XIV A), with its maximum length over 56 mm.—the inferior margin being broken—and its maximum breadth 24 mm. The suture bounding this line anteriorly is symmetrically disposed with regard to the left half of the coronal suture, which is normal, so the additional element appears to occupy part of the area normally occupied by the right parietal bone.

Other anatomical anomalies. The rarest anatomical anomaly noted in the whole series is a case of complete absence of the right auricular passage. This specimen (No. 324) is male and Plate XV A shows the condition of the affected region. There is no opening in the bone taking the place of the right auditory meatus. The corresponding region on the left side is normally formed. Apart from asymmetry in general form—the auricular region of the temporal and

adjoining parts of the base of the skull being more protruding on the left than on the right—there are no clear differences between the two sides of the base of the skull. Hrdlička (1932-3) has discussed similar cases of complete congenital absence of the external auditory meatus and tympanic bone, with reference to seven American specimens, all of which are affected on the right side only.

An orifice in the right temporal squama of a male skull (No. 266) was noted. This is the end of a canal leading through the bone to the interior of the brain cavity. Another male specimen (No. 301) exhibits complete blockage of the left jugular foramen (see Plate XV B), the right foramen being normal.

Wounds. Small healed wounds were found on several of the skulls, and they are noted in the remarks given in the appended tables of individual measurements. One male skull (No. 47) has a healed wound on the left malar bone, but all other injuries noted are on the cranial vault. The most severe examples of completely healed injury are on three male—No. 5 (frontal bone, Plate XVIII A), No. 190 (frontal bone), No. 301 (left parietal)—and four female—No. 419 (frontal, Plate XVIII C), No. 454 (frontal), No. 544 (frontal), and No. 670 (left parietal)—skulls. Only one specimen (No. 108, male, Plate III B), shows wounds which were probably inflicted not long before death. The outer table was completely removed in a region on the left, and another on the right, parietal: the edges of the affected area apparently show signs of healing, but the diploe is still exposed. There are long cracks in the vault of the same skull, but these are probably due to post-mortem injury.

Diseased and other conditions. I am indebted to Dr L. W. Proger and Dr A. M. El Batrawi for commenting on the following and some of the exceptional specimens previously described.

A female cranium (No. 662, Plate XVII) has what appears to be a diseased area extending over the greater part of the right side of the frontal bone. All the bones of the vault are softened and in a bad condition, with numerous small cracks, and the area in question is raised above the general level of the outer table. The roof of the right orbit is also affected and the endocranial surface of the right side of the frontal bone is slightly roughened. The inflammatory condition may possibly be due to osteomyelitis. Another male (No. 1, Plate XVIII B) has a roughened area on the parietal above the mid-point of the right side of the lambdoid suture. There seems to be no doubt that this is a pathological, and not a traumatic, lesion.

A cranium, presumed to be male (No. 382, Plate XVI) is peculiar on account of the "swollen" appearance of its brain-box. This suggests hydrocephaly, but the bones are not exceptionally thin. The basi-occipital is unusually short and broad, so the condition may be due to achondroplasia.

Artificially deformed skulls. There are eight of the Lachish skulls, all from Tomb 120, which are clearly artificially deformed, or which suggest this condition. The most marked cases are Nos. 381 (male) and 673 (female), both of

which happen to be metopic, and they are both young adult (see Plate VI). They are typical examples of fronto-occipital deformation, the two bones being flattened and a post-coronal depression being apparent in each case. Remarks on these two, and on the other six specimens, are in the appended tables of individual measurements. Of the latter, two male specimens (Nos. 376 and 378) show a clear suggestion of both frontal and occipital flattening, while the former shows slight post-coronal depression, but the latter does not. Nos. 377, 379, 380 (male) and 451 (female), show some degree of frontal flattening only, but the occipital bones appear to be perfectly normal: of these No. 379 is the only one showing any suggestion of a post-coronal depression. No. 375 (male) shows some degree of both frontal and occipital flattening but no post-coronal depression. Photographs of all these specimens, except No. 451, are reproduced in Plate VII. It is not possible to assert definitely that any of them were intentionally deformed in childhood. No. 380 has the coronal suture nearly obliterated and the sagittal and lambdoid open, and it is possible that its peculiar form is due solely to premature closing of the anterior suture. Artificial deformation was extremely rare in ancient Egypt, if it was ever practised there at all. There are records of affected specimens of earlier date than the Lachish from other parts of Western Asia, Crete, Cyprus, and some countries of Eastern Europe (see Dingwall, 1931).

Trepanned skulls. The three trepanned skulls have been described by Dr T. Wilson Parry (1936), and new photographs of them are given in Plates IV and V. He says that they are the first specimens exhibiting evidence of this surgical operation to have been found in Asia, and no others have been discovered since in the continent. On two of the skulls (No. 114, an ageing male, and No. 115, a young adult male) a quadrilateral of bone has been removed by sawing, and it is said that there is no evidence that the primitive operation was performed in the same way in any part of the world except Peru. The third skull (No. 340, a young adult male) shows the results of an operation of a different type. It is suggested that the individual had a depressed fracture, and that following this a piece of bone, which had become partly free as a result of the accident, was separated by sawing and removed. He survived long enough to enable the edges of the cavity to become completely healed, while the other two men must have died shortly after the operation. There appears to be no recorded case of a trepanned skull from ancient Egypt, Dr Batrawi (1935, Plate XV) has given a photograph of a young adult female specimen of Meroitic age from Lower Nubia with a circular opening on the right side of the frontal bone, supposed due to a trepan.

4. REMARKS ON THE CONDITION AND ANOMALIES OF THE JAWS OF THE LACHISH SKULLS

Considerably fewer than half of the skulls in the Lachish series have the upper dental arch complete or nearly complete. Table III gives statistics regarding the loss of teeth before death for the adult specimens having complete upper jaws, and for the smaller number of adult mandibles with the dental arch complete. The percentages of cases with no teeth lost before death are high, but it must be remembered that the age constitution, judging from the state of the sutures, indicates a younger group, on the average, than that expected in a cemetery population. The frequencies of adult upper jaws with one or both third molars absent are not unusual, but it is customary to find the female percentage greater than the male. The samples are too small, however, to give a reliable sexual comparison in this respect. Remarks on the condition and anomalies of the upper jaws, and of the few mandibles associated with crania, are given in the appended tables of individual measurements.

TABLE III

The frequencies of different conditions of the teeth in adult jaws having complete dental arcades

	Upper jaw		Lower jaw	
	♂	♀	♂	♀
(i) All teeth including third molars present at death	61	70	13	13
(ii) Third molars apparently absent and no teeth lost before death	6	8	1	4
(iii) One third molar apparently absent and no teeth lost before death	5	0	0	2
(iv) Third molars erupted, or believed erupted, and one or more teeth lost before death	55	29	12	12
(v) Third molars apparently absent, and one or more teeth lost before death	4	2	3	0
(vi) One third molar only apparently absent, and one or more teeth lost before death	1	1	1	1
Total no. of complete arches	132	110	30	32
Total no. with no teeth lost before death: (i), (ii) and (iii)	72 (54·5%)	78 (70·9%)	14 (46·7%)	19 (59·4%)
Total no. with one or both third molars believed unerupted: (ii), (iii), (v) and (vi)	16 (12·1%)	11 (10·0%)	5 (16·7%)	7 (21·9%)

Any dental anomalies of special interest were examined by Mr C. Bowdler Henry, M.R.C.S., and he has kindly allowed me to incorporate his notes in the following descriptions. The skiagrams reproduced in Plates XIX, XXII and XXIII were provided by him.

Deflected canines. There are four adult upper jaws with a canine on one side grossly deflected and not completely erupted.

No. 111, male. The unerupted right canine is misplaced and buried obliquely, so that the tip of the crown is situated in the palatal alveolus, between the sockets of the first and second incisors, and the apex of the root is seen (uncovered by bone) in the external surface of the maxilla over the apex of the socket of the first premolar. The third molars were congenitally absent, and all the other teeth were present and in good condition.

No. 142, male. The unerupted right canine is placed with the tip of its crown close to the anterior palatine fossa, and its root directed superiorly and posteriorly on the palatal side of the roots of the two premolars. It is possible that the left third molar was either abnormally small or else congenitally absent. The region of the right third molar is defective.

No. 401, female: Plate XXI D. The anterior part of the dental arch is damaged, and it is not possible to judge how many teeth were present. The unerupted left canine is placed with the tip of its crown in the anterior palatine fossa, and its root directed externally, posteriorly, and superiorly, towards the malar process of the maxilla, and placed above the apices of the premolars. The right third molar is normal, but it is probable that the left third molar was reduced in size.

No. 659, female. The left side of the palate and the part of the right side anterior to the canine are missing. The unerupted canine is placed with the tip of its crown near the normal position of the alveolus of the lateral incisor, and its root directed externally, posteriorly, and superiorly towards the malar process of the maxilla. The right third molar was probably erupted and lost before death.

Supernumerary denticles. There are two crania and one mandible exhibiting this anomaly.

No. 705, immature. The dental arch is symmetrical, with the left third molar erupting and the right apparently congenitally absent. There is a supernumerary denticle buried in the palate behind the right central incisor (see Plate XXI C). It is lying obliquely, with its crown in the anterior palatine fossa, and its tip impinging against the socket of the left central incisor. The condition of the residual root of the second right incisor suggests that this tooth had been broken during life.

No. 383, female. All the teeth appear to have been present and in good condition at death. There is a small supernumerary denticle placed unusually far back in the middle of the palate, to the left side of the suture (see Plate

XXII C, D). The skiagram shows that the apex of the crown is directed posteriorly.

No. 437, female. In the mandible all teeth were present at death, and there are five dental rudiments embedded in the outer alveolar margin. The positions of these can be seen from Plate XX. In the upper jaw all teeth were present at death, except the left third molar which had been lost. There are diastemae between the central incisors and between the lateral incisors and canines on both sides. There was post-normal occlusion of the jaws, and the teeth were markedly and irregularly worn.

Diastemae. There is one cranium showing diastemae between teeth, in addition to No. 437 described above.

No. 132, male. The third molars appear to have been congenitally absent, and all other teeth were present at death in the upper jaw. There is a diastema between the canine and first premolar on both sides (see Plate XXI B).

Misplaced and missing teeth, and retained milk teeth. In addition to the four examples of deflected canines described above, there are four crania and two mandibles falling in this category.

No. 72, male. The dental arch appears to be sufficiently roomy for the normal dentition, but crowding and some irregularity are present, due to the abnormal persistence of the right milk canine, and to the rotation of the left second premolar (see Plate XXIII C, D). The right third molar appears to have been congenitally absent, and the left is unerupted with the occlusal surface of its crown facing posteriorly and slightly laterally in the posterior wall of the antrum, and in the developmental position of the tooth. The deciduous canine is a well-developed tooth retained in its functional position. Limitation of space has caused the permanent right canine to be slightly rotated. The left second premolar is rotated so that its external, or buccal, cusp is antero-internal, and its internal, or palatal, cusp is postero-external.

No. 445, female. No teeth had been lost before death, but there are only sockets for three in the right side anterior to the second premolar (see Plate XXI A). It is probable that the lateral incisor was missing. The antero-inferior part of the inter-maxillary region is deflected to the right, and the anterior palatine foramen appears to be completely absent.

No. 496, female. The upper left third molar is below the alveolar margin and apparently impacted (see Plate XXII E).

No. 506, female. The dental arch is well formed, and all teeth except three molars are normally erupted. The left third molar was erupting at the time of death. The second and third right molars exhibit an unusual form of delayed eruption (see Plate XXII A, B). Neither tooth had emerged from the gum, although their direction and root formation appear to be normal, judging from the skiagram. The second molar is not impacted against the first, which might have accounted for the delayed eruption, but in fact they are clearly separated.

No. 1039, female. The left third molar of this mandible appears to be slightly "over-erupted", and it has a backward tilt. In the region of the second and third right molars, which are missing, there is a deep pathological excavation, suggestive of an abscess cavity or cyst. The place of the second premolar on the left side, which was congenitally absent, is taken by the retained deciduous second molar. The third molar on the left has an abnormally deep pit where the limbs of the cruciform fissures intersect: the aperture is 1 mm. in diameter. There are two smaller cavities in the occlusal surface of the second molar on the same side.

No. 1068, immature. This mandible had lost no teeth before death. The canine and first premolar on the right are unerupted and completely buried, the premolar lying obliquely towards the canine so that their crowns are in contact (see Plate XXIII A, B). A skiagram shows that the same teeth, unerupted on the other side, are in a similar condition.

Diseased conditions of the jaws. Only the more marked cases of diseases of the jaws were noted. Notes on a mandible (No. 1039), with a large abscess cavity or cyst, are given above, and there are three skulls and two mandibles exhibiting similar conditions.

No. 14, male. All upper teeth were present at death except the lateral incisor and canine on the right side. The shrunken and eburnated condition of the alveolus suggests that these two teeth might have been lost before death through traumatic injury.

No. 467, female. Nearly all teeth had been lost before death, and there is a large cyst in the anterior part of the palate, penetrating to the nasal aperture (see Plate XXIII F).

No. 469, female. All the upper molars had been lost before death, and there is a large abscess cavity or cyst in the molar region on the right side (see Plate XXIII E).

No. 1017, male. No teeth had been lost from the right side of this mandible before death, the left side being defective. There is a large carious cavity in the posterior half of the first molar, which evidently involved the dental pulp. In consequence of this, an abscess formed which discharged through a sinus in the external alveolar plate, over the apex of the posterior root of the tooth.

No. 1051, female. This mandible had lost no teeth before death, the third molars being congenitally absent. The bone of the external surface immediately above the left mental foramen is diseased.

Apparent adventitious filling of a tooth. A female cranium (No. 518) is remarkable on account of the fact that a piece of metal was found firmly embedded in, and level with, the occlusal surface of the second right molar near its centre. The fact that the surface of the metal that could be seen was flat, and that its edges conformed to the surface of the tooth, shows conclusively that it must have been where it was found during the life of the individual. Its appearance

was precisely similar to that of an artificial stopping. Plate XIX reproduces photographs of the whole palate and of the right molar region, before and after the removal of the metal, and a skiagram showing its depth. The filling was removed and its dimensions were found to be 1 × 3 mm. (exposed surface) by 1 mm. (height). The cavity in the tooth, which shows no sign of disease, is almost circular in form, and its position, where the fissures between the cusps join, is one in which a small pit is sometimes found. The description of mandible No. 1039, on p. 120 above, refers to pits of this kind in the occlusal surfaces of a second and third molar. There is a small pit in the surface of the second right molar of skull No. 518. These circumstances suggest that the person accidentally bit the piece of metal, which became lodged in the natural pit of the second molar and worn down as teeth are normally. Such an explanation appears much more plausible than an alternative one which might attribute the filling to surgical interference.

5. The statistical nature of the material

The statistical nature of the Lachish series is discussed in this section, topics considered being the question whether the male and female adult and immature samples can be supposed to represent the same population or not, sexual comparisons of average types and variabilities, and allied matters. As the series came from four tombs, which were adjoining, it is advisable to ask first whether there are any significant differences between the series from each, and whether it is legitimate to pool all the material for statistical purposes. In making all metrical comparisons, the specimens patently, or apparently, artificially deformed, those showing premature closing of the sagittal or coronal suture, and the one of unusual (? hydrocephalic) form were excluded. The numbers in these exceptional groups are in the table below, and measurements of the specimens, which are all adult, are given separately in the appended tables of individual measurements.

The numbers of anomalous crania

Tomb		♂	♀
120	Artificially deformed	6	1
	Premature closing of sagittal suture	11	6
	Premature closing of coronal suture	1	—
	Hydrocephalic?	1	—

When seen all together, the series from each tomb showed a variety of "types", but no specimens which clearly stood out from the others on account of their characters were noted, with the exception of two males, viz. No. 156, Tomb 120, and No. 179, Tomb 116. *Norma facialis* views of these two are

given in Plate III C, D, and it will be seen that they are of a very similar type, which is distinctly different from that of the male specimen (Plate X, right), which was selected because none of its measurements are at all peculiar (see p. 163 below). The cephalic index of No. 156 (81·7) is high, but not extreme, and that of No. 179 (78·4) is less exceptional. In spite of these two, it was thought best not to exclude any specimens from the series on account of their appearance.

In making male comparisons between tombs, it was necessary to pool data from Tombs 108 and 116, as the series from them are too small to stand alone. Male and female means are given in Table IV for various series from single tombs, or groups of tombs, in the case of those coefficient of racial likeness characters for which the pooled means for Tombs 107, 108 and 116 are based on more than ten skulls. It will be seen that the series from Tomb 120 is the only one long enough to provide adequate statistical constants. The others are just long

TABLE IV

Mean measurements of series of Lachish crania from different tombs

Character	Male series from tombs				Female series from tombs	
	120	107	108, 116	107, 108, 116	120	107, 108, 116
L	184·3 (266)	185·0 (32)	185·6 (24)	185·2 (56)	176·6 (200)	177·6 (59)
B	137·1 (276)	136·2 (28)	134·6 (23)	135·5 (51)	133·1 (204)	133·8 (57)
B'	95·4 (263)	96·5 (32)	95·1 (24)	95·9 (56)	92·2 (190)	92·1 (55)
H'	133·8 (226)	133·2 (24)	134·3 (18)	133·7 (42)	128·3 (166)	128·6 (47)
S	375·9 (221)	372·7 (20)	372·5 (14)	372·6 (34)	362·9 (164)	363·5 (48)
$\beta Q'$	308·7 (255)	307·7 (29)	308·3 (22)	308·0 (51)	297·6 (183)	298·8 (51)
U	518·1 (255)	519·0 (27)	517·1 (22)	518·2 (49)	500·1 (170)	501·4 (44)
fml	37·0 (211)	36·4 (21)	38·0 (15)	37·1 (36)	35·8 (151)	35·9 (42)
fmb	30·5 (213)	29·7 (20)	30·4 (11)	30·0 (31)	29·0 (154)	28·8 (38)
LB	100·6 (205)	101·3 (21)	101·1 (17)	101·2 (38)	96·6 (159)	95·7 (47)
$G'H$	70·1 (76)	68·8 (14)	72·5 (8)	70·2 (22)	67·2 (60)	66·4 (28)
NB	25·2 (96)	25·0 (18)	24·9 (9)	25·0 (27)	24·6 (62)	24·3 (26)
NH, L	51·3 (103)	51·7 (20)	52·1 (13)	51·8 (33)	49·1 (78)	48·6 (38)
O_1L	41·4 (117)	41·8 (21)	42·0 (10)	41·9 (31)	40·6 (78)	40·4 (33)
O_2L	32·8 (120)	33·6 (19)	33·3 (13)	33·5 (32)	33·1 (78)	33·4 (34)
G_1'	47·0 (72)	45·3 (16)	46·4 (12)	45·7 (28)	45·0 (63)	44·4 (27)
G_2	40·4 (43)	39·8 (8)	39·5 (6)	39·7 (14)	39·2 (48)	39·4 (20)
$100\ B/L$	74·5 (259)	73·6 (28)	72·6 (23)	73·1 (51)	75·5 (196)	75·4 (56)
$100\ H'/L$	72·7 (215)	72·5 (24)	72·0 (18)	72·3 (42)	72·7 (162)	72·8 (47)
$100\ B/H'$	102·5 (219)	103·2 (20)	100·8 (17)	102·1 (37)	103·7 (162)	104·3 (44)
$Oc.I.$	59·5 (238)	60·0 (25)	59·4 (15)	59·8 (40)	60·1 (166)	59·7 (49)
$100\ fmb/fml$	82·9 (195)	81·8 (16)	79·7 (11)	80·9 (27)	81·5 (134)	80·8 (33)
$100\ NB/NH$	49·6 (88)	48·7 (17)	47·9 (9)	48·4 (26)	50·3 (59)	49·7 (25)
$100\ O_2/O_1, L$	79·3 (113)	80·4 (18)	78·3 (10)	79·7 (28)	81·7 (69)	82·8 (31)
$100\ G_2/G_1'$	85·7 (33)	86·9 (6)	86·0 (5)	86·5 (11)	87·0 (39)	86·5 (16)
$N\angle$	64°·0 (70)	64°·9 (12)	62°·2 (7)	63°·9 (19)	64°·4 (54)	64°·7 (21)
$A\angle$	73°·8 (70)	74°·0 (12)	74°·5 (7)	74°·2 (19)	73°·9 (54)	73°·3 (21)

enough, however, to make comparison of interest. The following crude coefficients of racial likeness,* using standard deviations for the Tomb 120 series alone, were found:

Male, Tomb 120 (177·7†) and Tomb 107 (21·4) ... $-0{\cdot}45 \pm 0{\cdot}19$ (25‡).

Male, Tomb 120 (202·1) and Tombs 108, 116 (16·0) ... $0{\cdot}44 \pm 0{\cdot}21$ (20).

Male, Tomb 107 (23·9) and Tombs 108, 116 (16·6) ... $-0{\cdot}18 \pm 0{\cdot}21$ (20).

Male, Tomb 120 (167·3) and Tombs 107, 108, 116 (34·5) ... $0{\cdot}41 \pm 0{\cdot}18$ (27).

Female, Tomb 120 (116·3) and Tombs 107, 108, 116 (36·8) ... $-0{\cdot}20 \pm 0{\cdot}18$ (29).

It is known from experience that no reliance can be based on comparisons of measurements for short series of skulls. Two series, each being made up by fewer than twenty specimens, and actually representing two distinct races, are quite likely to indicate an insignificant difference, and series of fifty or more are usually required before any reliance can be placed on comparisons such as those given by coefficients of racial likeness. All of the five values above may be considered to differ insignificantly from zero, so the evidence, as far as it goes, indicates that the series from all the tombs may have been random samples from the same population. Larger samples would be required to justify this hypothesis in an adequate way, but for practical purposes there can be no objection to pooling all the material, on the supposition that it represents a single population. This conclusion is in conformity with the archaeological evidence.

It appeared worth while to compare the variabilities of two sub-groups of the total material. All the standard deviations given in Table V are for forty or more skulls. Two of the differences for corresponding constants might be considered just significant if considered by themselves (♂ L, $\Delta/\text{P.E. } \Delta = 3{\cdot}4$, and ♂ S_2, 3·2, the Tomb 120 constant being the greater in both these cases), but the conclusion for all characters must be that the two groups are almost identical in variability.

* With the usual notation, the form of the crude coefficient used is:

$$\frac{1}{M}\Sigma\left(\frac{n_s n_{s'}}{n_s+n_{s'}} \times \frac{(m_s - m_{s'})^2}{\sigma_s^2}\right) - 1 \pm 0{\cdot}67449\sqrt{\frac{2}{M}} = \frac{1}{M}\Sigma(\alpha) - 1 \pm 0{\cdot}67449\sqrt{\frac{2}{M}}.$$

If $\bar{n}_s$ is the mean number of skulls available for the characters used in the case of the first series, and $\bar{n}_{s'}$ is the same for the second series, then the "reduced" coefficient is defined to be:

$$50 \times \frac{\bar{n}_s + \bar{n}_{s'}}{\bar{n}_s \bar{n}_{s'}}\left\{\frac{1}{M}\Sigma(\alpha) - 1\right\} \pm 0{\cdot}67449\sqrt{\frac{2}{M}} \times 50 \times \frac{\bar{n}_s + \bar{n}_{s'}}{\bar{n}_s \bar{n}_{s'}}.$$

† This figure is the average number of skulls for the series ($\bar{n}$), in the case of the comparison in question, no means based on fewer than ten specimens being used.

‡ The number in brackets following a coefficient is the number of characters on which it is based.

TABLE V

Standard deviations for series from Tombs 107, 108 *and* 116 *combined, and from Tomb* 120

Character	Male		Female	
	Tomb 120	Tombs 107, 108, 116	Tomb 120	Tombs 107, 108, 116
L	6·03 ± 0·18*	4·82 ± 0·31	5·03 ± 0·17	5·19 ± 0·32
B	4·95 ± 0·14	5·65 ± 0·38	4·68 ± 0·16	3·95 ± 0·25
B'	4·16 ± 0·12	4·19 ± 0·27	4·53 ± 0·16	3·89 ± 0·25
H'	4·85 ± 0·15	5·73 ± 0·42	4·18 ± 0·15	4·32 ± 0·30
S_1	5·97 ± 0·18	5·84 ± 0·40	5·87 ± 0·20	5·13 ± 0·32
S_2	7·44 ± 0·21	5·96 ± 0·41	7·18 ± 0·25	7·02 ± 0·45
S_3	7·04 ± 0·22	5·89 ± 0·44	6·56 ± 0·24	6·84 ± 0·46
$\beta Q'$	9·54 ± 0·29	9·78 ± 0·65	9·47 ± 0·33	8·54 ± 0·57
U	13·86 ± 0·41	11·56 ± 0·79	12·54 ± 0·46	12·52 ± 0·40
100 B/L	3·02 ± 0·09	3·20 ± 0·22	3·01 ± 0·10	2·91 ± 0·18
100 H'/L	2·88 ± 0·09	3·18 ± 0·24	3·02 ± 0·10	2·51 ± 0·17
$Oc.I.$	2·62 ± 0·09	2·36 ± 0·18	2·51 ± 0·09	2·30 ± 0·16

* The symbol ± indicates probable errors throughout this paper. They were used instead of standard errors as probable errors have been given far more frequently than standard errors in craniometric studies.

The data for material from all four tombs were pooled, so that a single Lachish series, subdivided according to sex and age, is considered in all comparisons made below. It may be asked, next, whether the total adult male and female series can be supposed to represent the same population or not. The constants for them are given in Table VI.

The corresponding mean measurements of shape—indices and angles—for the two sexes are obviously very close. The differences only exceed three times their probable errors in the case of:

100 B/L (Δ/P.E. $\Delta = 6·7$), 100 B/H' (4·5), 100 $(B-H')/L$ (5·0), $Oc.I.$ (3·3),

100 O_2/O_1, L (6·5), 100 SS/SC (5·7), 100 fmb/fml (3·4), 100 S/C, ml (5·3).

In the case of the first six of these characters, it is normally found that the male and female means for long series show small differences of the same signs as those observed for the Lachish series. For the last character there is no good comparative material, but the longest series available, described by Woo (1937), shows a sex difference of the same sign. As far as can be told from these comparisons, the male and female Lachish series represent precisely the same population.

TABLE VI

Constants for the male and female series of adult crania from Lachish (all tombs combined)

Character.	Means (± P.E.)		Standard deviations		Coefficients of variation	
	Male	Female	Male	Female	Male	Female
C*	1425·1 (108)	1286·5 (89)	—	—	—	—
L	184·5 ± 0·22 (322)	176·8 ± 0·21 (259)	5·88 ± 0·16	5·09 ± 0·15	3·19 ± 0·08	2·88 ± 0·09
B	136·8 ± 0·19 (327)	133·3 ± 0·19 (261)	5·10 ± 0·13	4·55 ± 0·13	3·73 ± 0·10	3·41 ± 0·10
B'	95·5 ± 0·16 (319)	92·2 ± 0·19 (245)	4·26 ± 0·11	4·37 ± 0·13	4·46 ± 0·12	4·74 ± 0·14
H'	133·8 ± 0·20 (268)	128·4 ± 0·23 (213)	5·00 ± 0·15	5·05 ± 0·16	3·74 ± 0·11	3·93 ± 0·13
OH	115·1 ± 0·24 (108)	109·4 ± 0·30 (89)	3·69 ± 0·17	4·22 ± 0·21	3·20 ± 0·15	3·86 ± 0·20
S_1'	112·9 ± 0·18 (299)	108·7 ± 0·18 (248)	4·51 ± 0·12	4·11 ± 0·12	3·99 ± 0·11	3·78 ± 0·11
S_2'	116·0 ± 0·22 (323)	112·1 ± 0·23 (251)	5·88 ± 0·16	5·48 ± 0·16	5·07 ± 0·13	4·89 ± 0·15
S_3'	96·3 ± 0·19 (280)	94·0 ± 0·21 (216)	4·73 ± 0·13	4·52 ± 0·15	4·91 ± 0·14	4·81 ± 0·16
S_1	129·2 ± 0·24 (296)	124·7 ± 0·25 (248)	6·01 ± 0·17	5·75 ± 0·17	4·65 ± 0·13	4·61 ± 0·20
S_2	129·9 ± 0·27 (321)	124·9 ± 0·31 (249)	7·25 ± 0·19	7·16 ± 0·22	5·58 ± 0·15	5·73 ± 0·17
S_3	116·9 ± 0·28 (279)	113·3 ± 0·31 (215)	6·98 ± 0·20	6·68 ± 0·22	5·97 ± 0·17	5·89 ± 0·19
S	375·5 ± 0·54 (255)	363·0 ± 0·53 (212)	12·73 ± 0·38	11·48 ± 0·38	3·39 ± 0·10	3·16 ± 0·10
$\beta Q'$	308·6 ± 0·37 (306)	297·8 ± 0·41 (234)	9·60 ± 0·26	9·31 ± 0·29	3·11 ± 0·08	3·13 ± 0·10
U	518·1 ± 0·52 (304)	500·4 ± 0·58 (214)	13·51 ± 0·37	12·48 ± 0·41	2·61 ± 0·07	2·49 ± 0·08
fml	37·0 ± 0·11 (247)	35·8 ± 0·10 (193)	2·52 ± 0·08	2·14 ± 0·07	6·81 ± 0·21	5·98 ± 0·21
fmb	30·4 ± 0·09 (244)	28·9 ± 0·10 (192)	2·16 ± 0·07	1·99 ± 0·07	7·09 ± 0·22	6·87 ± 0·24
LB	100·7 ± 0·17 (243)	96·4 ± 0·20 (206)	3·82 ± 0·12	4·24 ± 0·14	3·79 ± 0·12	4·40 ± 0·15
GL	94·3 ± 0·34 (89)	90·6 ± 0·36 (76)	4·71 ± 0·24	4·65 ± 0·25	5·00 ± 0·25	5·13 ± 0·28
$G'H$	70·1 ± 0·30 (98)	66·9 ± 0·28 (87)	4·36 ± 0·21	3·94 ± 0·20	6·22 ± 0·30	5·88 ± 0·30
J	128·4 ± 0·48 (49)	121·3 ± 0·46 (40)	4·97 ± 0·34	4·28 ± 0·32	3·87 ± 0·26	3·53 ± 0·27
GB	94·4 ± 0·31 (107)	91·8 ± 0·30 (81)	4·81 ± 0·22	4·04 ± 0·21	5·09 ± 0·24	4·40 ± 0·23
NB	25·2 ± 0·11 (123)	24·5 ± 0·12 (88)	1·74 ± 0·07	1·73 ± 0·09	6·89 ± 0·30	7·49 ± 0·38
NH, L	51·4 ± 0·15 (136)	48·9 ± 0·17 (116)	2·64 ± 0·11	2·70 ± 0·12	5·14 ± 0·21	5·52 ± 0·25
SC	10·5 ± 0·09 (188)	10·0 ± 0·10 (146)	1·80 ± 0·06	1·87 ± 0·07	17·14 ± 0·61	18·63 ± 0·76
SS	5·0 ± 0·06 (186)	4·5 ± 0·07 (134)	1·23 ± 0·04	1·21 ± 0·05	24·40 ± 0·90	27·07 ± 1·19
O_2L	32·9 ± 0·11 (152)	33·2 ± 0·13 (112)	2·05 ± 0·08	2·01 ± 0·09	6·22 ± 0·24	6·05 ± 0·27
O_1L	41·5 ± 0·09 (148)	40·6 ± 0·11 (111)	1·64 ± 0·06	1·65 ± 0·07	3·95 ± 0·15	4·07 ± 0·19
G_1'	46·7 ± 0·19 (100)	44·8 ± 0·21 (90)	2·79 ± 0·13	3·00 ± 0·15	5·98 ± 0·29	6·69 ± 0·34
G_2	40·3 ± 0·20 (57)	39·2 ± 0·19 (68)	2·29 ± 0·14	2·28 ± 0·13	5·69 ± 0·36	5·81 ± 0·34
Ml_1	59·8 ± 0·33 (81)	56·3 ± 0·36 (58)	4·44 ± 0·24	4·11 ± 0·26	7·43 ± 0·40	7·29 ± 0·46
Ml_2	49·8 ± 0·20 (134)	47·2 ± 0·19 (101)	3·35 ± 0·14	2·83 ± 0·13	6·73 ± 0·28	5·99 ± 0·29
$C(ml)$	52·4 ± 0·25 (83)	49·9 ± 0·32 (58)	3·32 ± 0·17	3·56 ± 0·22	6·33 ± 0·33	7·13 ± 0·45
$S(ml)$	11·9 ± 0·12 (83)	10·6 ± 0·13 (58)	1·63 ± 0·09	1·47 ± 0·09	13·72 ± 0·73	13·85 ± 0·88
$100\ B/L$	74·3 ± 0·12 (310)	75·5 ± 0·13 (252)	3·08 ± 0·08	2·99 ± 0·09	—	—
$100\ H'/L$	72·7 ± 0·12 (257)	72·7 ± 0·14 (209)	2·93 ± 0·08	2·92 ± 0·10	—	—
$100\ B/H'$	102·4 ± 0·20 (256)	103·8 ± 0·24 (206)	4·72 ± 0·14	5·11 ± 0·17	—	—
$100\ (B-H')/L$	1·7 ± 0·14 (246)	2·8 ± 0·17 (202)	3·25 ± 0·10	3·52 ± 0·12	—	—
$Or.I.$	59·5 ± 0·10 (278)	60·0 ± 0·11 (215)	2·59 ± 0·07	2·48 ± 0·08	—	—
$100\ fmb/fml$	82·7 ± 0·26 (222)	81·4 ± 0·28 (167)	5·82 ± 0·19	5·31 ± 0·20	—	—
$100\ G'H/GB$	74·6 ± 0·39 (63)	72·9 ± 0·45 (59)	4·64 ± 0·28	5·14 ± 0·32	—	—
$100\ NB/NH, L$	49·4 ± 0·25 (114)	50·2 ± 0·28 (84)	4·00 ± 0·18	3·81 ± 0·20	—	—
$100\ O_2/O_1, L$	79·4 ± 0·28 (141)	82·0 ± 0·30 (100)	4·96 ± 0·20	4·50 ± 0·21	—	—
$100\ G_2/G_1'$	85·9 ± 0·50 (44)	86·9 ± 0·52 (55)	4·95 ± 0·36	5·73 ± 0·37	—	—
$100\ SS/SC$	48·6 ± 0·48 (186)	44·2 ± 0·59 (134)	9·71 ± 0·34	10·17 ± 0·42	—	—
$100\ Ml_2/Ml_1$	84·0 ± 0·55 (78)	84·5 ± 0·64 (57)	7·16 ± 0·39	7·19 ± 0·45	—	—
$100\ S/C(ml)$	22·8 ± 0·20 (83)	21·2 ± 0·22 (58)	2·72 ± 0·14	2·49 ± 0·16	—	—
$P\angle$	86°·0 ± 0·24 (81)	84°·9 ± 0·28 (62)	3°·14 ± 0·17	3°·23 ± 0·20	—	—
$N\angle$	64°·0 ± 0·24 (89)	64°·5 ± 0·30 (75)	3°·43 ± 0·17	3°·79 ± 0·21	—	—
$A\angle$	73°·9 ± 0·20 (89)	73°·7 ± 0·27 (75)	2°·83 ± 0·14	3°·50 ± 0·19	—	—
$B\angle$	42°·0 ± 0·20 (89)	41°·8 ± 0·23 (75)	2°·77 ± 0·14	2°·90 ± 0·16	—	—

* Determined from reconstruction formulae: see Appendix I.

The few juvenile and more numerous adolescent skulls together, a total of sixty-one unsexed specimens, give the mean indices in the following table, where comparison is made with the male and female adult means.

	Male	*Female*	*Immature*
100 B/L	74·3 ± 0·12 (310)	75·5 ± 0·13 (252)	76·2 (53)
100 H'/L	72·7 ± 0·12 (257)	72·7 ± 0·14 (209)	72·6 (41)
100 B/H'	102·4 ± 0·20 (256)	103·8 ± 0·24 (206)	104·0 (39)
100 $(B-H')/L$	1·7 ± 0·14 (246)	2·8 ± 0·17 (202)	2·7 (38)
100 fmb/fml	82·7 ± 0·26 (222)	81·4 ± 0·28 (167)	76·5 (35)
100 NB/NH	49·4 ± 0·25 (114)	50·2 ± 0·28 (84)	52·0 (22)
100 O_2/O_1, L	79·4 ± 0·28 (141)	82·0 ± 0·30 (100)	83·8 (22)

The sequence, male adult—female adult—immature (unsexed) mean, is expected in the case of the last three of these indices. The first four give means for the immature series which differ quite insignificantly from the female adult means, and a close correspondence of this kind is expected. There is thus every reason to believe that the children and adults belonged to the same population.

The sex ratios (male mean/female mean) for a few measurements of size are given in Table VII for the Lachish, three Egyptian, and four English series. All the means involved are based on eighty-six or more crania.

TABLE VII

*Sex ratios for the Lachish and other series**

	Lachish	Egyptian, E	Egyptian, Kerma	Egyptian, Denderah	English, Farringdon St	English, Whitechapel	English, Spitalfields	English, Hythe
B	1·027	1·025	1·029	1·031	1·049	1·045	1·050	1·046
S	1·034	1·034	1·039	—	1·046	1·039	1·051	1·040
U	1·035	1·038	1·044	—	1·044	1·041	–	1·038
B'	1·036	1·027	1·040	—	1·038	1·053	1·014	1·041
H'	1·042	1·038†	1·044	1·041	1·059	1·059	1·053	1·054
L	1·043	1·047	1·045	1·044	1·040	1·048	1·052	1·038
LB	1·045	1·053	1·054	1·047	1·045	1·066	1·058	1·053

* The series are: Egyptian E, Gizeh, 26th–30th dynasties (Davin & Pearson, 1924); Kerma, 12th–13th dynasties (Collett, 1933); Denderah, 6th–12th dynasties (Thomson & MacIver, 1905); Farringdon St, seventeenth century (Hooke, 1926); Whitechapel, seventeenth century (Macdonell, 1904); Spitalfields, Roman or medieval (Morant & Hoadley, 1931); Hythe, medieval (Stoessiger & Morant, 1932).

† For H in place of H'. These very similar calvarial heights give almost identical sex ratios in the case of long series for which both are given.

The characters are arranged in the table in order of the sex ratios for the Lachish series. The three Egyptian series give very similar orders, that for the

Kerma being particularly like the Lachish. As is usually found, the preponderance of the average male over the average female skull is clearly different for different measurements, the sex ratios being rather less for cranial diameters than for stature and the lengths of the long bones. It is curious that the orders in which the characters are arranged by the Lachish and Egyptian sex ratios are distinctly different from all those given by the constants for the four English series, while these last show a general, though not very close, correspondence *inter se*. There appears to be a distinct difference between the sex differences of the Lachish and ancient Egyptian races, on the one hand, and the later English ones, on the other.

Measures of variability for the total male and female adult series of crania from Lachish are given in Table VI. Comparison of the relative degrees of variation exhibited by the material for the two sexes may be considered first. For the thirty-three absolute measurements the male standard deviation exceeds the corresponding female value in twenty-four cases, and for the remaining nine the position is reversed. In one or two instances the excess of the male over the female constant is clearly significant, but where the female standard deviation exceeds the male the differences are quite insignificant. In comparing parallel samples representing the same population it is customary to find a slight preponderance of male over female variation for size characters. A closer approach to equality is normally found for these if coefficients of variation are used.

The following comparisons are based on coefficients of variation for absolute (size) characters, and standard deviations for measurements of shape (indices and angles), the total number of characters being fifty. In twenty-four cases the male measure of variation exceeds the female, and for the other twenty-six the position is reversed. The only difference which exceeds three times its probable error is for LB (Δ/P.E. $\Delta = 3{\cdot}2$), but no importance whatever can be attached to a divergence of this order in fifty comparisons. The male and female series thus show a remarkably close approach to equality in variation, and the agreement in this respect again suggests that a single population is represented.

It may be asked next how the variability of this population compares with that of others which are usually accepted as being racially homogeneous. It will be sufficient to make comparisons with the long series of 26th–30th dynasty skulls from Gizeh (Davin & Pearson, 1924), as these have often been used for the purpose. There are thirty-three characters for which the constants required (coefficients of variation for absolute measurements and standard deviations for indices and angles) are available. In the case of the male series the Lachish value exceeds the Egyptian E in fifteen cases, and the Egyptian E exceeds the Lachish in the remaining eighteen. Differences exceeding three times their probable errors are only found in the case of three measurements of shape, viz. $100\, B/L$ (Δ/P.E. $\Delta = 4{\cdot}0$, Lachish σ the greater), $A\angle$ (4·1, Egyptian E greater),

and *Oc.I.* (8·3, Egyptian *E* greater). In the case of the female series the Lachish constant exceeds the Egyptian *E* in twenty-three cases, and the Egyptian *E* exceeds the Lachish in the remaining ten. The "significant" differences are for $B\angle$ (3·1), B' (3·8), 100 B/L (4·2) and LB (4·6), the Lachish variability being the greater for these four, and *Oc.I.* (7·2), in which case the Egyptian $E\sigma$ exceeds the Lachish. It is clear that there is a close agreement between the variabilities of the two populations, and, in fact, it is not possible to say that one was more homogeneous than the other. No importance can be attached to any of the differences for single characters, except those for the occipital index, which is decidedly more variable for the Egyptian series in the case of both sexes.

Comparison with data for other series shows that the Lachish standard deviations of the occipital index are quite unexceptional, while those for the Egyptian *E* series are peculiarly large, in view of the fact that all other characters for it indicate a low order of variability. It may be noted that a few of the Lachish crania (excluded from the series) are clearly artificially deformed. If artificial deformation had been generally practised to a slight degree, this would have been expected to increase the variability of the occipital index. Such an effect is not observed, however, and it is very probable that the exceptional cases noted are the only ones exhibiting artificial deformation.

All the statistical evidence thus points to the fact that the male and female adult and immature Lachish series represent random samples from a single population, which showed the same order of variability as ancient Egyptian populations. It is known that these were slightly less variable than Western European populations in later times.

6. The relationships of the Lachish and Ancient Egyptian populations judged from comparisons of mean measurements

The Lachish is the first series of skulls of any period from Palestine which is of an adequate length, and for which adequate measurements are available. A rough comparison of mean measurements for it showed that the type is very similar to that of certain ancient Egyptian series, and this was fully substantiated by statistical comparisons. As far as can be judged from the appearance of the skulls, no clear distinction from dynastic Egyptian material is evident. Comparisons with Egyptian series only are dealt with in this section, and reference is made to other material from Palestine in § 9.

Far more skulls have been preserved and studied from an anthropological point of view from Egypt than from any other country in the world. The vast majority of these relate to Roman and earlier times, and no comparisons with the meagre post-Roman material from the country were made. Most of the

...n and allied series of male skulls*

	Kerma	Abydos	Sheikh Abd
	12th–13th dyn.	18th dyn.	18th dyn.
	Collett	Thomson & MacIver	Thomson & MacIver
	107·3	40·9	40·0
	13	14	15
	7·80 (29)	1·29 ± 0·32 (14)	3·06 (14)
	7·77 (31)	25·41 (14)	7·91 (14)
	5·05 (14)	21·24 (14)	5·73 (14)
	1·11 ± 0·21 (31)	13·05 (14)	2·50 ± 0·37 (14)
	1·15 ± 0·19 (14)	11·59 (14)	3·19 (14)
	3·84 (14)	8·44 (14)	9·36 (14)
	15·52 (31)	2·08 ± 0·64 (14)	18·48 (14)
	10·56 (14)	1·96 ± 0·58 (14)	11·36 (14)
	5·85 (14)	11·08 (14)	4·44 (14)
	16·14 (31)	13·12 (14)	15·70 (14)
	6·27 (28)	14·84 (14)	4·67 (14)
	5·75 (14)	7·08 (14)	0·22 ± 0·41 (14)
		10·17 (14)	7·33 (14)
	10·17 (14)		10·97 (14)
	7·33 (14)	10·97 (14)	
	14·06 (29)	11·97 (14)	21·08 (14)
	8·09 (29)	2·87 ± 0·49 (14)	8·40 (14)
	7·75 (30)	8·92 (14)	6·62 (14)
	15·25 (31)	6·23 (14)	13·02 (14)
	7·09 (14)	1·47 ± 0·42 (14)	6·05 (14)
	13·82 (14)	5·89 (14)	7·70 (14)
	5·01 (28)	12·95 (14)	4·30 (14)

...of characters. A probable error is only given if a coefficient is

longer series of male Egyptian skulls representing populations of the period considered have been compared statistically by Morant (1925). A selection of the material with which he dealt was made, and certain series were excluded for various reasons, viz.:

(i) Fouquet's predynastic—the so-called Aeneolithic—series from Middle Egypt. The male means of size characters for this series are decidedly larger than those for any other predynastic series dealt with in Morant's paper, or described since, while its mean indices are very close to those for some contemporary series. It appears to be very probable that the distinction was due, either to a process of selection which favoured the larger specimens, or to inaccurate sexing. While the series is suspect it appears better not to use it for comparative purposes.

(ii) The series measured by Broca and Chantre, for which the means are unsatisfactory as they are not recorded in fine enough units.

(iii) Four other series, each comprising fewer than thirty skulls.

(iv) The El Kubanieh South series, measured by Toldt, which relates to a period covering Early and Middle Dynastic times, while all the other series relate to shorter intervals.

The modern Abyssinian series measured by Sergi, and reduced by Morant, is included in our list. This accounts for eighteen of the series used. The three other ancient Egyptian series included are the 9th dynasty from Sedment (Woo, 1930); the Kerma (Collett, 1933)—from a Nubian site, although the population was unquestionably of Egyptian type—and the pooled early predynastic Badari series (Morant, 1935), measured by Stoessiger (1927) and Derry. Including the Lachish, there is thus a total of twenty-two male series, made up by about 3000 skulls. The localities from which the collections were obtained are shown in the sketch-map, Fig. 2.

Reduced coefficients of racial likeness between all pairs of the twenty-two series are given in Table VIII. Many of these were obtained from the crude coefficients given by Morant (1925), and others were copied from tables given by Woo (1930), and Collett (1933). In addition to these I was able to use a number of unpublished values calculated by Morant, and to complete the table I calculated all the coefficients with the Lachish series and several others which were wanting. The $\bar{n}$ for a particular series in the table gives the average number of skulls on which the means are based, in the case of all the characters used in computing the coefficients which are available for the series. The $\bar{n}$'s range from 29·9 to 855·4, and there are only six, including the Lachish, greater than 100. Experience has suggested that a series of fifty complete skulls of one sex, or a larger number of incomplete specimens providing an equivalent amount of evidence, is required in order to give reliable racial comparisons with

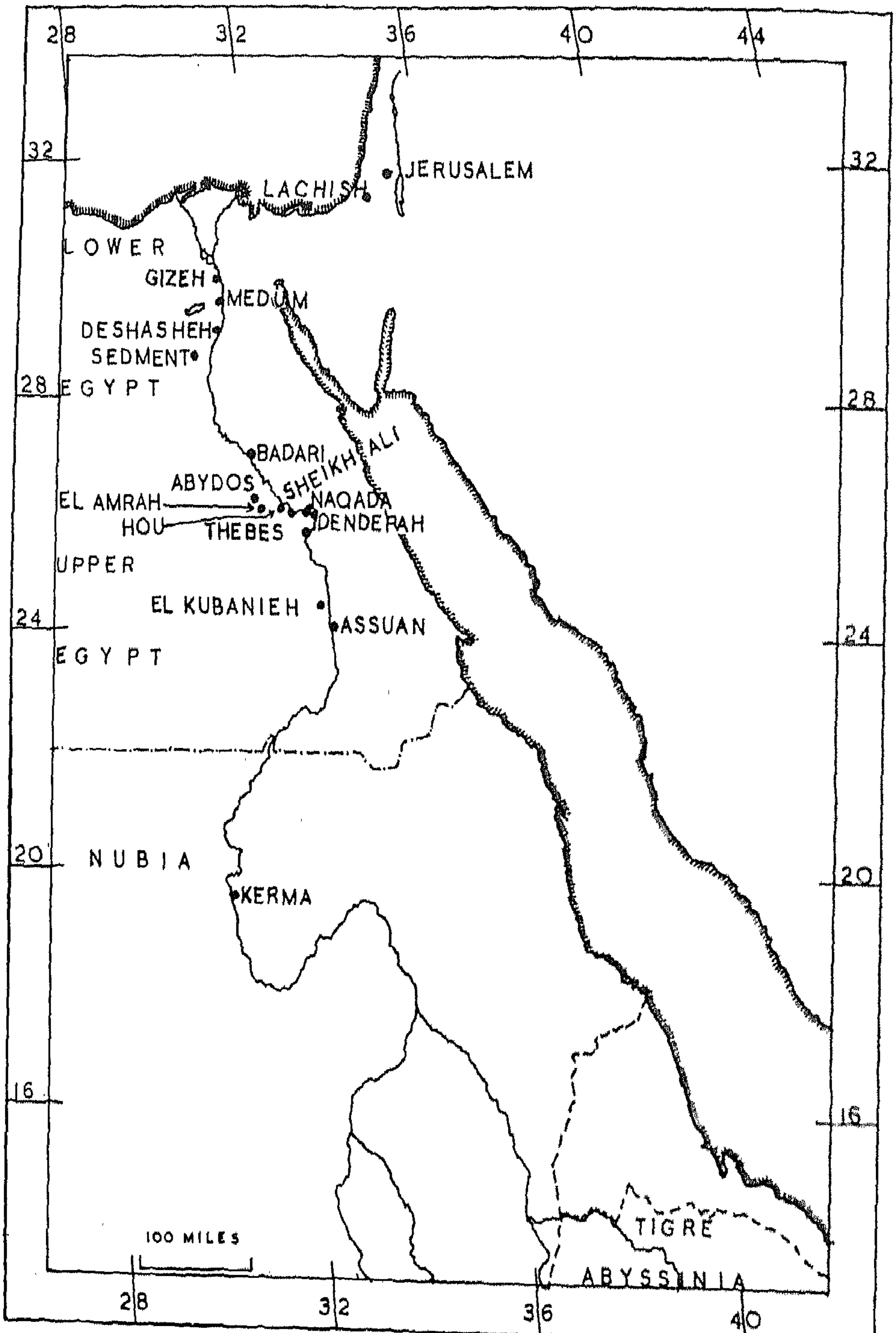


Fig. 2. Sketch-map of Egypt and neighbouring countries showing localities from which cranial series were obtained.

a particular population. Of the twenty-two series, there are eight which fail to satisfy this ideal requirement, but these may be supposed large enough to give fairly reliable comparisons.

When possible coefficients of racial likeness between cranial series have been based on a standard set of thirty-one characters, nineteen of these being absolute measurements (chords and arcs), seven indices, and three angles. For the Egyptian material considered it was only possible to use the complete set in a few cases, owing to the fact that one or several of the measurements are not available for the majority of the series. In all cases as many as possible of the selected characters were used in computing the coefficients. The distribution of the numbers of characters on which the 231 coefficients could be based is:

No. of characters	14	24	25	26	27	28	29	30	31
No. of comparisons	165	2	5	5	5	16	14	7	12

More than half of the coefficients are thus based on only fourteen characters, all these being comparisons of the ten series measured by Thomson and MacIver with one another, and with the remaining twelve series. The coefficients between these twelve are all based on a number of characters which can be supposed adequate, but it may be suggested that comparisons between them and those for which only fourteen characters are involved are likely to be unreliable. The question of how far the generalized estimate of resemblance for fourteen characters is likely to diverge from that obtained from about twice as many was investigated by computing some of the more reliable coefficients for the incomplete set of fourteen characters available for Thomson and MacIver's series. Comparisons of the corresponding reduced coefficients for (*a*) all of the characters available, and (*b*) for the fourteen characters only, are made in Table IX, in the case of fifteen pairs of series selected at random.

For seven of the fifteen comparisons the reduced coefficient for fourteen characters is greater than the corresponding value for the larger number, and markedly greater in four cases; for the remaining eight comparisons the position is reversed, but there are no marked differences between the pairs. It appears that the generalized estimate of divergence derived from the fourteen characters can usually be expected to give a fairly close approximation to that which would be obtained from about twice as many characters, but occasionally it will suggest a rather misleading conclusion. In the case of the material considered, the fourteen characters show a tendency to indicate a wider separation of the types than that likely to be found if more evidence were available, and this might have been anticipated from the fact that they include the characters which normally show the most significant differences in the comparisons of Ancient Egyptian material, viz. B, $100\ B/L$, J and $N\angle$.

TABLE IX

Comparisons of corresponding pairs of reduced coefficients of racial likeness (male series) for all coefficient characters available and for a set of fourteen characters

	Reduced coefficients of racial likeness	
	All characters	14 characters
Lachish (191·5*) with Thebes, 18th–20th dyn. (54·1)	2·13 ± 0·23 (25)	0·01 ± 0·30
Naqada *A* and *Q* (64·9) with El Kubanieh North (33·1)	3·91 ± 0·41 (28)	6·41 ± 0·58
Abydos, 18th–19th dyn. (31·5) with Thebes, 18th–20th dyn. (54·1)	5·13 ± 0·48 (26)	8·76 ± 0·64
Lachish (191·5) with Thebes, 18th–21st dyn. (167·9)	5·27 ± 0·11 (26)	3·23 ± 0·14
Lachish (191·5) with Gizeh, 26th–30th dyn. (869·5)	5·77 ± 0·06 (29)	5·37 ± 0·08
Lachish (191·5) with Abyssinian (62·6)	6·89 ± 0·20 (26)	4·25 ± 0·27
Lachish (191·5) with Kerma (107·6)	7·86 ± 0·13 (29)	7·78 ± 0·19
Abyssinian (62·6) with El Kubanieh North (33·1)	8·08 ± 0·44 (25)	5·56 ± 0·59
Sedment (37·5) with Thebes, 18th–21st dyn. (167·9)	9·18 ± 0·29 (30)	11·09 ± 0·42
Thebes, 18th–20th dyn. (54·1) with Abydos, 1st dyn. Royal (33·2)	12·03 ± 0·44 (29)	10·63 ± 0·62
Abyssinian (62·6) with Abydos, 18th–19th dyn. (31·5)	13·04 ± 0·46 (25)	19·37 ± 0·61
Sedment (37·5) with Abydos, 18th–19th dyn. (31·5)	13·12 ± 0·60 (28)	13·97 ± 0·73
Naqada *A* and *Q* (64·9) with Gizeh, 26th–30th dyn. (869·5)	14·35 ± 0·14 (30)	19·03 ± 0·21
Lachish (191·5) with Badari (61·1)	24·12 ± 0·22 (29)	21·07 ± 0·27
Badari (61·1) with Abydos, 1st dyn. Royal (33·2)	36·24 ± 0·44 (31)	45·33 ± 0·59

* The numbers in brackets following the titles of the series give the average numbers of skulls ($\bar{n}$'s) in the case of the fourteen coefficient of racial likeness character available for all the series. In the case of the comparisons for "all characters" the $\bar{n}$'s are changed, but they are invariably of the same order. Fuller particulars regarding the series are given in Table VIII.

The point in question can be examined in an indirect way by comparing the distributions of reduced coefficients in Table VIII based (*a*) on fourteen characters, and (*b*) on twenty-four to thirty-one characters. These are:

	Insignificant*	Significant and <3·5	3·5–5	5–13	13–20	20–30	30–40	40–50	Totals
14 characters	9	16	18	72	27	17	5	1	165
24–31 characters	0	4	6	31	17	5	2	1	66
All comparisons	9	20	24	103	44	22	7	2	231

* A coefficient is counted here as insignificant if it differs from zero by less than 3·5 times its probable error.

Allowing for the differences in their sizes, these two distributions appear to be very similar, and it is clear that the coefficients for fourteen characters only do not show any marked tendency to diverge from the others on the average.

It is curious that there are no insignificant coefficients based on twenty-four or more characters, but little importance need be attached to this fact.

In computing all the reduced coefficients of racial likeness given in Table VIII, the Egyptian *E* standard deviations (Davin & Pearson, 1924) were used. It is thus assumed that the variabilities of all the populations are so closely similar that any one of them can be adequately represented by the constants for the longest series of the group available. It has been shown above that the Lachish standard deviations are very similar to the corresponding values for the *E* series. The constants for three of the other series—viz. the Badari, Sedment, and Kerma—have been compared with the Egyptian *E* in the same way, and it was found that the variabilities indicated by them are also of almost precisely the same order. At the same time it is known that the Ancient Egyptian series show a clear tendency to be rather less variable than Western European series of later times.

In the comparison of any two series the full coefficient of racial likeness involves both sets of standard deviations. The question of the extent to which the assumption that the population variabilities were precisely the same is likely to affect reduced coefficients was examined in an indirect way in the case of a few comparisons. For twenty-nine of the comparisons reduced coefficients happened to be available computed by using both (*a*) the Egyptian *E* σ's, and (*b*) the Farringdon St (seventeenth-century London, Hooke, 1926) σ's. In every case the former value is greater than the latter, as would have been anticipated, and the ratio of the reduced coefficients found with Egyptian *E* σ's to the corresponding values found with Farringdon St σ's, range from 1·05 to 1·79, while for 21 of the 29 ratios, the range is between 1·3 and 1·6. The Farringdon St variablities are obviously inappropriate for use in comparisons of Ancient Egyptian material.

For the ten cases giving the largest ratios the reduced coefficient was also computed using the Lachish σ's, and the values for the three sets are given in Table X. The ratios of coefficients found with the Egyptian *E* σ's to the corresponding values found with the Lachish σ's range from 0·91 to 1·17. For seven of the ten comparisons the use of the Lachish σ's gives the lower value, but all the differences here are so small that they may be considered of no importance.

The reduced coefficients of racial likeness which were used in this paper to provide a classification of a number of Ancient Egyptian and related populations are clearly not statistical estimates of an ideal kind. Owing to the nature of the material, certain devices have to be used in practice which are likely to have an appreciable affect on the coefficients obtained. It has been shown above that there is good reason to believe that the use of a constant set of standard deviations does not distort the value to any appreciable extent. The use of different sets of the complete group of thirty-one characters appears to be of far more practical importance. More than half of the 231 coefficients computed are

actually based on fourteen characters only, the numbers for the remainder ranging from twenty-four to thirty-one. Comparisons with the shorter series might have been omitted, but this would have reduced the evidence to such an extent that it appeared better to include them, while remembering their imperfection. There are other objections to the constants. It is known that the theoretical requirement that all the characters used should be uncorrelated with one another

TABLE X

Corresponding reduced coefficients of racial likeness for male cranial series computed by using (1) *the Egyptian E*, (2) *the Lachish, and* (3) *the Farringdon St standard deviations*

	No. of characters	With Egyptian *E* σ's	With Lachish σ's	With Farringdon St. σ's
Abyssinians with Abydos (Early predyn.) ($\bar{n}$ 62·6) ($\bar{n}$ 40·9)	14	12·00 ± 0·52	10·76 ± 0·52	7·64 ± 0·52
Abyssinians with El Amrah and Hou (Late predyn.) ($\bar{n}$ 62·6) ($\bar{n}$ 105·6)	14	7·04 ± 0·32	6·33 ± 0·32	4·61 ± 0·32
El Amrah and Hou (Late predyn.) with Thebes (18–20 dyn.) ($\bar{n}$ 105·6) ($\bar{n}$ 54·1)	14	10·08 ± 0·36	8·86 ± 0·36	6·61 ± 0·36
Abydos (Early predyn.) with Abydos (12–15 dyn.) ($\bar{n}$ 40·9) ($\bar{n}$ 65·9)	14	4·99 ± 0·51	4·34 ± 0·51	3·13 ± 0·51
Naqada *A* and *Q* with Abydos (12–15 dyn.) ($\bar{n}$ 64·9) ($\bar{n}$ 65·9)	14	2·38 ± 0·39	2·39 ± 0·39	1·33 ± 0·39
Thebes (18–20 dyn.) with Thebes (18–21 dyn.) ($\bar{n}$ 53·1) ($\bar{n}$ 166·4)	23	1·43 ± 0·24	1·57 ± 0·24	0·90 ± 0·24
Naqada *A* and *Q* with Thebes (18–20 dyn.) ($\bar{n}$ 66·9) ($\bar{n}$ 53·4)	27	8·27 ± 0·30	7·08 ± 0·30	5·18 ± 0·30
Thebes (18–20 dyn.) with Egyptian *E* ($\bar{n}$ 53·2) ($\bar{n}$ 866·0)	25	3·37 ± 0·18	3·45 ± 0·18	2·35 ± 0·18
Denderah (Roman) with Abydos (12–15 dyn.) ($\bar{n}$ 49·3) ($\bar{n}$ 65·9)	14	6·81 ± 0·45	5·88 ± 0·45	4·34 ± 0·45
Thebes (18–20 dyn.) with Deshasheh (4–5 dyn.) ($\bar{n}$ 54·1) ($\bar{n}$ 39·9)	14	2·39 ± 0·56	2·25 ± 0·56	1·42 ± 0·56

is not fulfilled, some pairs of the characters being almost certainly quite highly correlated in all the series, but it is probable that the relative values of the estimates of resemblance are not more disturbed by this than by some of the other factors. In the case of a particular measurement used, comparisons are only made between mean readings said to have been obtained by following a particular definition of the measurement. Even so, the personal equations of different measurers may have been large enough, in some cases, to disturb the results quite appreciably. It is not possible to investigate this matter in the case of all the material used, but it must be recognized that the disturbance due to it is probably quite sufficient to invalidate any rigid application of statistical rules to the material. Mistakes in sexing the crania provide another possible source of error, though probably this is one of minor importance.

It is not at all likely, of course, that the extent to which a particular coefficient is distorted is determined by the accumulation of the various kinds of disturbing factors mentioned. One may hope that they will usually counteract one another. In view of all the circumstances, it is obvious that no great reliance can be placed on the numerical accuracy of any particular reduced coefficient of racial likeness, or on the comparison of any pair of the constants. It would be most inadvisable, for example, to compare the difference between any pair with regard to the probable error of this difference. It is not unreasonable, however, to accept the coefficients as indicating different, rather broad, grades of resemblance, and it is in this sense that they are used for purposes of classification here, in conformity with previous biometric practice.

The classification of the series treated in this section is based on reduced coefficients of racial likeness for pairs of the male series: the female series are in most cases shorter, but they may be expected to lead to very similar conclusions. Corresponding male and female values for the Lachish compared with six other series are given in Table XI. For the first of these comparisons the difference

TABLE XI

*Corresponding male and female reduced coefficients of racial likeness for comparisons of the Lachish with other series**

		Abydos 18th dyn.	Thebes 18th–20th dyn.	Abydos and Hou 12th–15th dyn.
		♂ (40·0) ♀ (66·1)	♂ (53·8) ♀ (43·5)	♂ (65·9) ♀ (87·4)
Lachish	♂ (184·0) ♀ (147·1)	1·29 ± 0·32 (14) 1·05 ± 0·28 (14)	2·13 ± 0·23 (24) 6·37 ± 0·29 (24)	4·62 ± 0·26 (14) 2·51 ± 0·23 (14)

		Egyptian *E* 26th–30th dyn.	Denderah 6th–12th dyn.	Kerma 12th–13th dyn.
		♂ (859·5) ♀ (569·7)	♂ (168·6) ♀ (140·4)	♂ (107·5) ♀ (84·2)
Lachish	♂ (184·0) ♀ (147·1)	5·74 ± 0·06 (29) 6·87 ± 0·08 (29)	6·57 ± 0·14 (14) 8·67 ± 0·17 (14)	7·86 ± 0·13 (29) 12·66 ± 0·17 (29)

* The series compared in this table with the Lachish, are ones for which fuller particulars are given in Table VIII. The numbers in brackets after the sex signs are the $\bar{n}$'s for the numbers of skulls on which the means used in calculating the coefficients are based. The $\bar{n}$'s given for the Lachish series are for the twenty-nine characters on which two of the comparisons are based, and its values for the other groups of characters are very close to these.

between the male and female coefficient is of no account, but for some of the others the difference is appreciable, even in the case of the Lachish compared with the Egyptian *E* series, for which all the means involved are based on an adequate number of skulls. The estimates of resemblance are in fact rather different for the male and female comparisons, though there is a fairly close agreement in their orders, except in the case of the comparison with the 18th–20th dynasty series from Thebes. The means for this series suggest that the groups of male and female skulls composing it may not have represented precisely the same population, and any slight discordance of this kind is likely to have an appreciable effect on the comparisons considered. Errors in sexing are also likely to affect them to some extent. We are again led to the conclusion that there is no justification for attaching any importance to small differences between coefficients of racial likeness.

The conclusions regarding the racial affinities of the populations represented by the twenty-two Ancient Egyptian and allied series may now be considered. Reduced coefficients of racial likeness between all pairs of them are given in Table VIII, and the arrangement suggested by the lowest orders of these constants is shown in Fig. 3. Three sets of series, representing races in other parts of the world, have been treated in the same way: the results for a North American set have been published by von Bonin & Morant (1938), for an Asiatic set by Woo & Morant (1932), and for a European set by Morant (1928). For this last material crude coefficients of racial likeness only are given for forty-one series, four of which relate to North African peoples. I am indebted to Morant for the corresponding reduced coefficients, which he has calculated but which have not been published. The four North African series were excluded, so the numerical results to be considered refer to thirty-seven European series.

It is of interest to compare the distributions of the reduced coefficients for the new Egyptian and the three other sets. These are given in Table XII. There are marked differences between the ranges of the constants for the four sets of material. On the average the Egyptian shows much lower values than the other three sets, the highest reduced coefficient for it being 47·5, while for each of the others more than one quarter of the coefficients exceed 30 and several values greater than 100 are found.

The relationships of the modern races of man are of such a kind that they can be roughly divided into groups, so that all the members of a particular group show close resemblances, while there are links between the different groups. This conception of constellations of races, as it were, which are joined to one another is undoubtedly useful, though it is often difficult to give precision to it, owing to the existence of subgroups within the major groups. The distributions in Table XII suggest that the selected set of twenty-two Ancient Egyptian and allied series may be considered to represent a single constellation of populations which had very similar types, while each of the other three sets

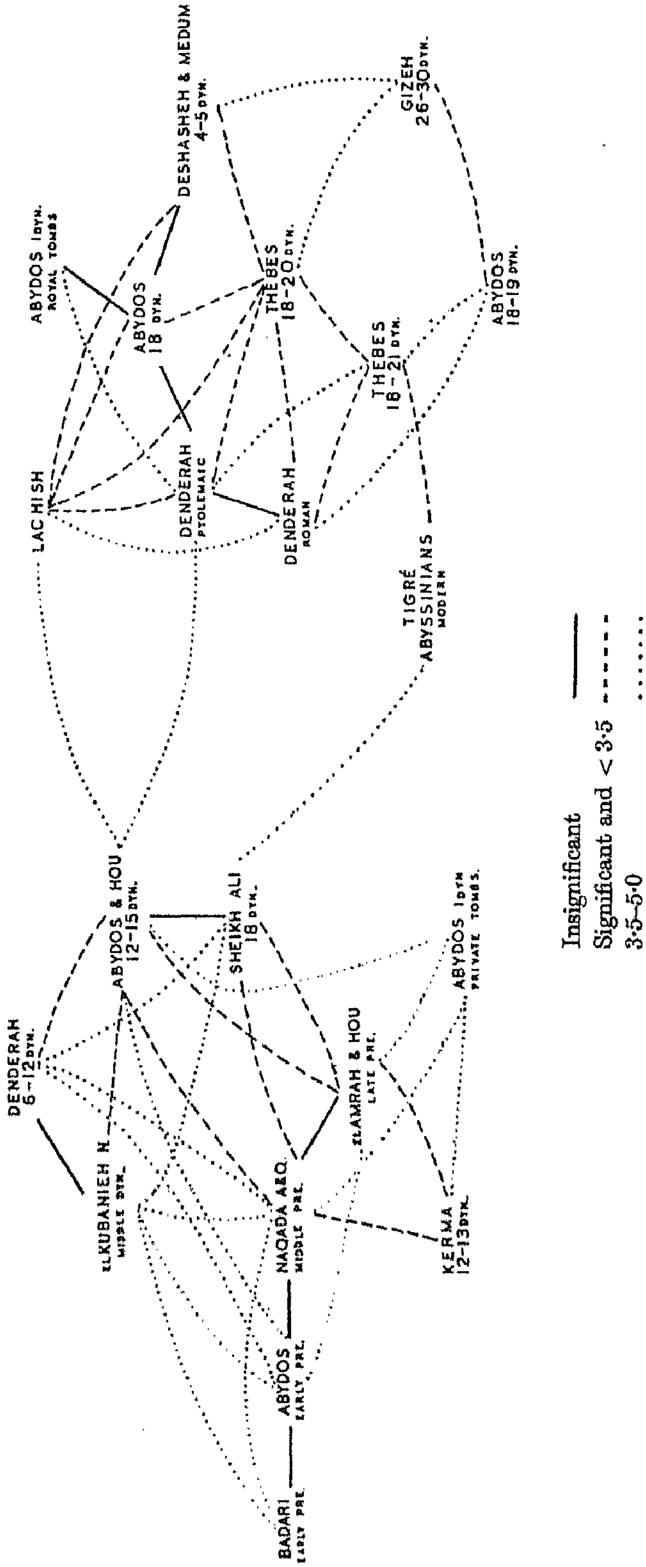


Fig. 3. Reduced coefficients of racial likeness between Ancient Egyptian and allied series of male skulls.

represents two or more constellations. The Egyptian series actually relate to peoples who lived from early predynastic to modern times, a span of about 7000 years, but there is no suggestion that the group changed radically within this period either as the results of evolutionary change, or owing to infiltration by alien peoples. The area represented extends from Abyssinia to Palestine, though nearly all the material comes from Egypt.

TABLE XII

Distributions of reduced coefficients of racial likeness for various groups of male cranial series

Groups of series	No. of series	Coefficients of racial likeness							Total comparisons
		<5	5–13	13–50	50–100	100–200	200–400	400–430	
Ancient Egyptian and allied	22	53 (22·9%)	103 (44·6%)	75 (32·5%)	...	...	.	.	231
European	37	33 (5·0%)	102 (15·3%)	362 (54·4%)	135 (20·3%)	34 (5·1%)	..	..	666
North American Indian	16	2 (1·7%)	9 (7·5%)	59 (49·2%)	45 (37·5%)	5 (4·2%)	.	..	120
Asiatic	26	7 (2·2%)	23 (7·1%)	102 (31·4%)	110 (33·8%)	56 (17·2%)	26 (8·0%)	1 (0·3%)	325

It is known from experience that in the cross comparisons of two groups of cranial series, representing two distinct families of races (A and B), some lower reduced coefficients will be found between them than *some* for pairs of series belonging to group A, or than *some* for pairs of series belonging to group B. In comparisons of the Ancient Egyptian with North American Indian series, for example, it would be anticipated that a certain proportion of the coefficients would be below 50, though the majority would doubtless exceed this limit. The peculiarity of the generalized measure of resemblance employed must be supposed due to the peculiar nature of the material, and it need not be taken to indicate any defect in the method. A classification based on such generalized estimates of resemblance must evidently proceed on certain lines, and the way in which the constants can most usefully be employed for the purpose has to be determined empirically.

Some of the cross comparisons $A \times B$ referred to give lower reduced coefficients than some for pairs of A's, or some for pairs of B's, but at the same time they have always been found to exceed a certain limit, so that the lowest order of coefficients are not represented by any $A \times B$ comparisons. This suggests that any classification of the material ought to be derived solely from coefficients less than the limit in question, i.e. from the evidence of close resemblance only, such as is never found, as far as experience goes, between any two populations belonging to distinct families of races.

The earlier studies of groups of cranial series treated by the method of the coefficient of racial likeness suggested that the limit in question might be taken as a reduced coefficient of about 20, so that the classification would be based on all values less than 20, while all greater than this limit would be neglected entirely. Recent experience of the cross comparisons between North American Indian and Asiatic series (von Bonin & Morant, 1938) has shown that it is safer to employ a considerably lower limiting value, which was provisionally taken at 13. If the particular set of material treated is made up by a considerable number of very closely related series, a clearer picture of the interrelationships is obtained, however, by taking a still lower limiting value.

In the present case the limit was chosen arbitrarily at 5·0, that is to say the evidence presented by all reduced coefficients less than 5·0 is taken into account —bringing in 23 % of the total of 231 comparisons—while all values greater than 5·0 are neglected. It should be realized that a reduced coefficient less than 5·0 represents very close similarity of the cranial types compared, such as has hitherto only been found in cases where close relationship between the populations represented was anticipated, from the cultural (historical or archaeological) evidence.

The grades of resemblance indicated by reduced coefficients of racial likeness less than 5·0 are shown in Fig. 3. The full lines represent "insignificant" coefficients—i.e. values which differ from zero by less than 3·5 times their probable errors. In these cases the two series might have represented precisely the same population, as far as can be seen from the direct comparison, though, owing to the imperfection of the material, no stress can be laid on the difference between an "insignificant" coefficient and one of the same order which indicates clear differentiation judged by the ratio of the constant to its probable error. The second grade of resemblance is represented by coefficients which differ significantly from zero, but which are less than 3·5, and the third by values between 3·5 and 5·0.

It should be stressed that any coefficients of racial likeness less than 5·0 indicate a very close resemblance of the cranial types compared. There are two adequately long male series of seventeenth-century London skulls, from Whitechapel and Farringdon St, and the coefficient between them is 3·8 (see Morant & Hoadley, 1931, p. 233), so a value of this order can only be taken to show the degree of divergence which may be found between parochial subgroups of the same population.

The arrangement shown in Fig. 3 was arrived at by placing the series, as far as possible, in position relative to one another so that their distances apart are proportional to the estimates of resemblance considered. One of the series treated—viz. that of the 9th dynasty skulls from Sedment—is not shown, because the lowest reduced coefficient found with it is 5·1 (with the 4th–5th dynasty series from Deshasheh and Medum).

There is a clear suggestion of a division into two groups.

Group A includes eight series from sites in Upper Egypt (covering a stretch of the Nile, about 100 miles in length, from Badari in the north to Thebes in the south—see Fig. 2), a series from El Kubanieh (100 miles south of Thebes), and a Nubian series from Kerma. The period represented by these ten series extends from early predynastic times to the 18th dynasty.

Group B includes seven series from the same area in Upper Egypt, two from Lower Egypt, one from Abyssinia, and the Lachish from Palestine. The Upper Egyptian series here range in time from the 1st dynasty—only represented by a single series of skulls from Royal Tombs—to Roman times. The two Lower Egyptian series and the Lachish (of about 25th dynasty date) fall within the same period, and the Abyssinian series is of modern date.

The population of Upper Egypt is represented by fifteen of the twenty-two series, representing a time sequence from early predynastic to Roman times. All the series of earlier date than the 18th dynasty except one fall in group *A*, and show very close interconnexions. At the same time there is a clear suggestion that the type was changing gradually with time, the earliest predynastic (Badari) series being at one extreme, and the middle dynastic series at the other. The only aberrant Upper Egyptian series of a date prior to the 18th dynasty is one from Royal Tombs at Abydos, and it may be supposed that these represent an intrusive group coming from Upper Egypt. The fact that the early Upper Egyptian type persisted almost unchanged until the 18th dynasty is evidenced by the series of this date from Sheikh Ali.

There are four Upper Egyptian series of the 18th–21st dynasties—two from Abydos and two from Thebes—which are quite distinct from the Upper Egyptian (*A*) group of series, and which must hence be supposed to represent mainly an intrusive population in the region. Immediately before the 18th dynasty the whole of Egypt was under the Hyksos dominion, and when this lapsed at the end of the 17th dynasty the country was in an unsettled state. It may be noted that if there were immigrants into Upper Egypt at this time, they were most likely to have been of the ruling classes, who would have settled and been buried in the principal towns in Upper Egypt, viz. Abydos and Thebes. The population of Upper Egypt in later times is only represented by two series from Denderah, one of Ptolemaic and the other of Roman date. The types of these two are very similar to those of the presumed intrusive groups of the 18th–21st dynasties, though they stand somewhat closer to the indigenous Upper Egyptian types. This may obviously be explained as being due to some slight degree of intermixture between the intrusive group and the settled population of Upper Egypt between the 18th dynasty and Ptolemaic times.

If an explanation of this kind is on the right lines, then it becomes necessary to discover the source of the people who are presumed to have migrated to Upper Egypt about the time of the 18th dynasty. The relationships of the

cranial series in Fig. 3 provide a clear answer to this question, as they show intimate connexions between two late Upper Egyptian series and two of the three Lower Egyptian series available, viz. the 4th–5th dynasty series from Deshasheh and Medum, and the 26th–30th dynasty from Gizeh, forming with them what is here called the *B* group.

Unfortunately, the evidence available is quite inadequate to provide an outline of the racial history of Lower Egypt. A third series from that region, viz. one from Sedment of 9th dynasty skulls, is not shown in the diagram, because its lowest reduced coefficient exceeds 5·0; this is 5·1 with the 4th and 5th dynasty series from Deshasheh and Medum, sites close to Sedment. The three Lower Egyptian series are very similar to one another, and they stand on the same side, as it were, of all the Upper Egyptian material. Another factor which has to be taken into account in interpreting the evidence is that the Upper Egyptian type was apparently becoming gradually modified, *prior* to the 18th dynasty, in the direction of the Lower Egyptian type. The evidence suggests the following general conclusions:

(1) It may be supposed that originally the populations of Upper and Lower Egypt formed two distinct groups, the purest representatives of these known being the Early predynastic from Badari, and the 4th–5th dynasty series from Deshasheh and Medum. These are the earliest series available from Upper and Lower Egypt, respectively. For convenience, the groups may be called, following Morant, the Upper (corresponding to our Group *A*), and the Lower (Group *B*) Egyptian types. From the earliest times until about the end of the 17th dynasty, the type of the Upper Egyptian population became *gradually* modified in the direction of that of the Lower Egyptian. This may be supposed due to a gradual infiltration of Lower Egyptians into Upper Egypt. The relationships of the Kerma series from Nubia are of particular interest in this connexion. It is of 12th–13th dynasty date, but its closest connexions are with two predynastic series from Upper Egypt. Hence it may be supposed that the Kerma people represent the descendants of colonists who left Upper Egypt in predynastic times, and that this stock remained stable, and was not modified by intermixture, though the parent group itself was changing owing to contacts with the North.

There is only one Upper Egyptian series of earlier date than the 18th dynasty which stands apart from the constellation formed by all the others. This is of 1st dynasty skulls from Royal Tombs at Abydos, and as their type is very similar to the Lower Egyptian, it may be supposed that they represent an intrusive group from that region. This group, which was probably small, may have been absorbed in the Upper Egyptian population without affecting the type of the latter appreciably.

(2) The situation in Upper Egypt appears to have changed radically in the 18th dynasty. There is one Upper Egyptian series of this date (from Sheikh Ali)

which is very similar to the earlier series from that region, but four others (ranging from the 18th to the 21st dynasty) stand quite apart and show close relationships with the Lower Egyptian series. The movement from Lower to Upper Egypt appears to have been greatly accelerated about the time of the 18th dynasty. The evidence suggests that at this time it virtually became a peaceful invasion of Upper Egypt, which resulted in the almost complete displacement of the earlier population there. This view is supported by the fact that there are no series whatever later than the 18th dynasty of the early Upper Egyptian type, while two late series from Denderah are still of Lower Egyptian type, though somewhat closer to the early Upper Egyptian than are the 18th–21st dynasty series from Thebes and Abydos. These relationships suggest that the prevailing population in Upper Egypt after the 18th dynasty was of Lower Egyptian origin, and that it became mixed to some extent with descendants of the earlier population of the region.

The fact that a modern population from the north of Abyssinia stands between the two Ancient Egyptian groups is interesting, but far more evidence would obviously be required to elucidate the significance of this relationship.

The relationships of the Lachish series may now be considered. All its closest connexions are with series of the Lower Egyptian type, and these are close enough to suggest that the population in the Palestinian town was entirely, or almost completely, of Egyptian origin. The skeletal material from Lachish is believed to represent people who died about 700 B.C., which is the time of the 25th dynasty in Egypt. The series shows closest resemblance, however, not to the 26th–30th dynasty series from Gizeh, but to the early dynastic series from Lower Egypt (Deshasheh and Medum), and to three Upper Egyptian series of 18th dynasty or later dates which are assumed to represent populations of Lower Egyptian origin. There is also a rather less close connexion between the Lachish and one of the Upper Egyptian types (Abydos and Hou, 12th–15th dynasty). These relationships suggest that the Lachish population represents descendants of a colonizing group of men and women, which was derived primarily from Upper Egypt, at some time later than the 18th dynasty, and which maintained its type unchanged—free from intermixture—until 700 B.C. or later It is not at all unlikely, of course, that a colonial group in Palestine, such as the one which settled in Lachish, originally included people from several parts of Egypt, and not from a single restricted locality. The characteristics of the Lachish sample give no hint of heterogeneity, however, which may be due to the fact that any diversity which originally existed in the group was obscured by intermarriage within the community for several generations. The evidence suggests quite clearly that the Lachish people were derived principally from a population of Upper Egypt which was itself derived principally from emigrants who left Lower Egypt about the time of the 18th dynasty.

7. Comparisons of the Lachish and Ancient Egyptian series for characters considered singly

Comparison of the Lachish and a number of Ancient Egyptian and allied cranial series, based on the averages for a number of characters considered in conjunction, have been made in the preceding section. The means for the same material will now be compared for characters considered singly.

A convenient way of summarizing the statistical comparisons of means for groups having a number of features recorded, is provided by considering the α's calculated in computing the coefficients of racial likeness. An α is an approximation to the square of a quantity which is the difference of the means divided by the standard error of this difference. If an α is greater than 10, it may be supposed that the difference between the two means is clearly significant. The percentages of α's greater than 10 may be used to distinguish characters which are practically constant for all the groups from those which frequently denote differentiation.

The twenty-one comparisons between the Lachish and each of the other series may be considered first. The numbers of α's available for this set range for the different characters from seven to twenty-one, only fourteen of the thirty-one coefficient of racial likeness characters being available for all the series.* These fourteen may be considered first, and they give the following grouping:

(a) characters showing a high proportion of significant differences:

B (percentage of α's$>10=47{\cdot}6$), $100\,B/L$ (42·9), and $100\,B/H'$ (42·9);

(b) characters showing a lesser proportion of significant differences:

$N\angle$ (28·6), H' (23·8), NH (23·8), J (19·0), and $100\,H'/L$ (19·0);

(c) characters showing few significant differences:

L (14·3), $100\,NB/NH$ (9·5), $A\angle$ (9·5), $G'H$ (4·8), NB (4·8), and LB (0·0).

It is clear that some characters can be supposed practically constant for all the series, while others show many significant differences.

The remaining characters which can be treated in the same way are only available for numbers varying from eight to twelve of the twenty-two series. The numbers of comparisons are very restricted for these, but they suggest that S, O_2, G'_1, $100\,G'H/GB$, $100\,G_2/G'_1$, and $P\angle$ are practically constant for all the series, while the Lachish is most frequently differentiated by B', *Oc.I.*, *fml*, *fmb*, U, and rather less frequently by G_2, O_1, $100\,O_2/O_1$ and $100\,fmb/fml$.

Table XIII gives the ranges of certain male means for the Upper Egyptian group of series (i.e. the ten on the left-hand side of Fig. 3), the means for the Lachish

* The transverse arc ($\beta Q'$) for the Lachish series is not available for any of the others, and the estimated cranial capacities were not included, so the maximum number of characters used in computing coefficients with the Lachish series is 29.

alone, and the ranges for the remaining nine comprising the Lower Egyptian group (i.e. all on the right-hand side of Fig. 3 except the Lachish and Abyssinian). There is seen to be complete separation of the ranges for the two sets of series, with the Lachish falling within the Lower Egyptian limits, in the case of B, B', U, and 100 B/H', and a close approach to the same condition in the case of J and 100 B/L. All these measurements are transverse breadths, or indices including a breadth, and there is no doubt that the values of the reduced coefficients of racial likeness are largely determined by differences between them. Owing to the use of correlated measurements, the differences in cranial breadth affect the generalized measure of resemblance unduly. This factor was apparently showing far more significant variation in Ancient Egyptian populations than any other relating to the cranium.

TABLE XIII

Ranges of mean measurements for two groups of Ancient Egyptian male skulls and the Lachish means

Series	Period	B	J	B'
Upper Egyptian type	Early predyn.–18th dyn.	131·4–134·3 (10)*	123·6–127·5 (8)	90·4–92·8 (4)
Lachish	*ca.* 25th dyn.	136·8	128·4	95·5
Lower Egyptian type	1st dyn.–Roman	135·3–139·3 (9)	127·5–131·3 (8)	93·0–96·2 (5)

Series	Period	U	100 B/L	100 B/H'
Upper Egyptian type	Early predyn.–18th dyn.	500·0–510·4 (4)	71·7–73·7 (10)	98·1–101·1 (10)
Lachish	*ca.* 25th dyn.	518·1	74·3	102·4
Lower Egyptian type	1st dyn.–Roman	510·8–518·7 (5)	73·7–76·0 (9)	102·3–106·4 (9)

Series	Period	L	H'	$N\angle'$
Upper Egyptian type	Early predyn.–18th dyn.	182·2–185·2 (10)	132·4–135·9 (10)	64°·9–67°·0 (8)
Lachish	*ca.* 25th dyn.	184·6	133·8	64°·0
Lower Egyptian type	1st dyn.–Roman	181·4–185·8 (9)	130·7–136·0 (9)	63°·4–66°·1 (8)

Series	Period	*fml*	*fmb*	*Oc.I.*
Upper Egyptian type	Early predyn.–18th dyn.	34·9–36·4 (4)	28·7–30·1 (4)	60·2–62·3 (4)
Lachish	*ca.* 25th dyn.	37·0	30·5	59·5
Lower Egyptian type	1st dyn.–Roman	35·1–37·0 (5)	29·7–30·2 (5)	59·9–61·5 (3)

* The numbers in brackets indicate the number of series to which the ranges relate.

The remaining characters treated in the table make no clear distinctions between the two sets of series, and the same situation is observed in the case of all the other coefficient of racial likeness characters. The Upper Egyptian series show a slight tendency to be more prognathous than the Lower Egyptian (judging by $N\angle$), but the two ranges overlap appreciably. It may be noted that the foraminal length and breadth of the Lachish type are extremely large, and its occipital index is extremely low compared with those for the other series.

The majority of the measurements not used in computing the coefficients appear to be fairly constant for all the Ancient Egyptian series, and the Lachish means fall within their ranges. This is so far the bimaxillary breadth (GB), in spite of the clear differences found between the bizygomatic and calvarial breadths. The index 100 (B-H')/L does make a clear distinction between the Upper and Lower Egyptian sets of series, but this again appears to be due to the fact that a breadth measurement is involved.

The simotic measurements, giving estimates of the "flattening" of the nasal bridge, are only available for a few series. Comparison with the data for these given by Woo & Morant (1934), shows that the breadth of the nasal bones (SC) is unexceptional for the Lachish type, but the subtense (SS) and index (100 SS/SC) are decidedly larger for it than for the Badari, Kerma, Sedment, and an Ancient Nubian type. The greater curvature of the nasal bones in the Lachish skulls places them within the range found for European populations.

Comparative material for the malar bone measurements is still more restricted. Comparison with the data given by Woo (1937) suggests that the Lachish means are unexceptional, but for the index measuring the curvature of a horizontal section of the bone (100 S/C, ml). For this the Palestine series has a mean which is decidedly greater than those for the two Egyptian series, and this mean places it close to the extreme for the types from all parts of the world hitherto described. For the vast majority of measurements the Lachish skulls are not distinguishable as a group from series representing Ancient Egyptian populations. The most distinguishing features of the type appear to be the exceptional curvature of the nasal bridge and horizontal sections of the malar bones. This divergence is suggestive, but more abundant comparative material would be needed to assess its significance.

8. THE CONTOURS OF THE LACHISH SERIES

Contours of adult skulls in the Lachish series were drawn in accordance with the methods used in a number of earlier craniometric studies published in *Biometrika*, and a selection of the total group had to be made for this purpose. Types were constructed from the measurements of these contours in the usual way (Figs. 4–9), and they relate to material from the four tombs combined (see p. 123). The selection was made by first excluding all specimens too incomplete to give the Frankfurt orientation, and then excluding from the

remainder those for which the horizontal and transverse sections are defective to more than a slight extent. The totals included are 108 male and 89 female crania, and almost all the measurements used in constructing the types (Tables XV–XVII) are available for every one of these specimens, except in the case of facial measurements of the sagittal contours for which the numbers vary appreciably. The type contours of the Lachish series are based on samples which are almost as large as any previously used for the purpose.

A comparison is made in Table XIV between certain measurements of the sagittal type contours and corresponding caliper measurements for the total

TABLE XIV

A comparison of mean caliper and type contour measurements

Character	Male		Female	
	Contour	Caliper	Contour	Caliper
L	185·3 (108)	184·5 (322)	178·8 (89)	176·8 (259)
H'	135·0 (96)	133·8 (268)	127·6 (74)	128·4 (213)
S_1'	113·6 (108)	112·9 (299)	110·2 (89)	108·7 (248)
S_2'	116·1 (108)	116·0 (323)	111·8 (89)	112·1 (251)
S_3'	96·7 (97)	96·3 (280)	94·5 (79)	94·0 (216)
fml	36·7 (96)	37·0 (247)	35·1 (74)	35·8 (193)
$G'H$	69·7 (66)	70·1 (98)	66·2 (44)	66·8 (87)
GL	94·2 (62)	94·3 (89)	91·1 (38)	90·6 (76)
LB	101·8 (96)	100·7 (243)	97·1 (74)	96·4 (206)
$P\angle$	87°·0 (66)	86°·0 (81)	86°·5 (44)	84°·9 (62)
$N\angle$	63°·4 (62)	64°·0 (89)	65°·1 (38)	64°·5 (75)
$A\angle$	75°·2 (62)	73°·9 (89)	73°·6 (38)	73°·7 (75)
$B\angle$	41°·4 (62)	42°·0 (89)	41°·3 (38)	41°·8 (75)

series. The former are either used in the construction of the type ($G'H$, GL, LB), or calculated from measurements used in its construction (H', S_1', S_2', S_3', fml, $P\angle$, $N\angle$, $A\angle$ and $B\angle$), or measured on the figure (L). In comparing the means of the two kinds, it must not be forgotten that one series represented is a selected group of the other. It might be anticipated that the process of selection described would favour the larger and stronger skulls, more likely to be well preserved, and hence rather larger means would be expected for the contour series. The divergences between corresponding values are actually found to be very small. Of the nine absolute measurements the male contour mean exceeds the caliper in six cases, the largest difference being 1·2 mm., and the position is reversed for the other three, which show a maximum difference of 0·4 mm. For the same measurements the female contour mean exceeds the caliper value in five instances, the maximum difference being 2·0 mm., and the maximum difference for chords showing the caliper greater than the contour mean is

0·8 mm. The angles show a close correspondence. On the whole there is a good agreement between the measurements obtained in the two ways, suggesting that the contours were drawn with sufficient accuracy, and that the sub-series gives a fair representation of the type for the total series.

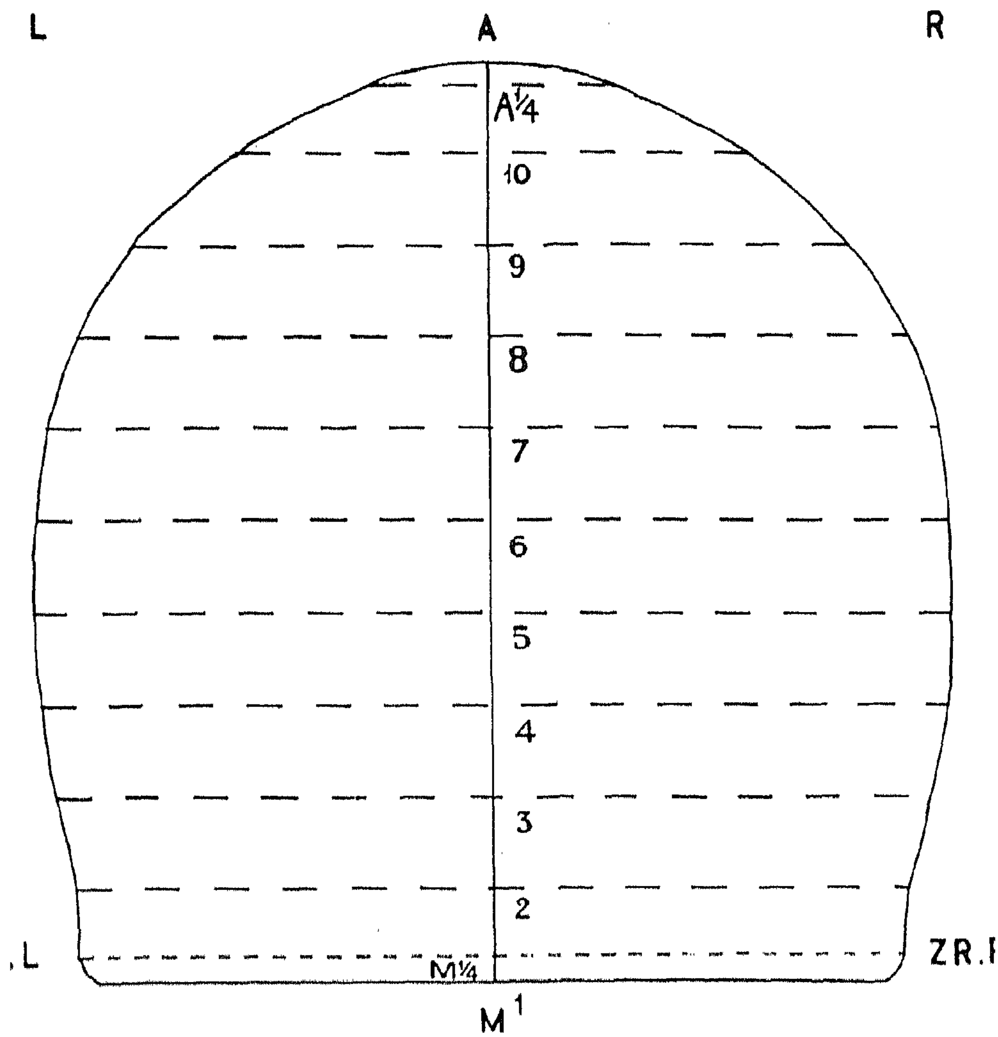


Fig. 4. Transverse type contour based on 108 male Lachish skulls.

The type contours (Figs. 4–9) have no striking peculiarities, and superficially, at any rate, they appear to be very similar to those given for Ancient Egyptian and even some European series. They are average in size, show an orthognathous facial skeleton, and only moderate muscular development. The male contours are larger than the female in all respects to the same extent as is usually found.

Fig. 10 shows the three male Lachish types and those given by Miss Collett (1933) for the Kerma, and by Miss Stoessiger (1927) for the Badari series, superposed. There is clearly a close agreement between the average contours for these two Ancient Egyptian and the Palestinian series, though the differences are probably almost as great as those which would be found between any pairs of the types of Ancient Egyptian series. The Kerma and Badari series are assigned

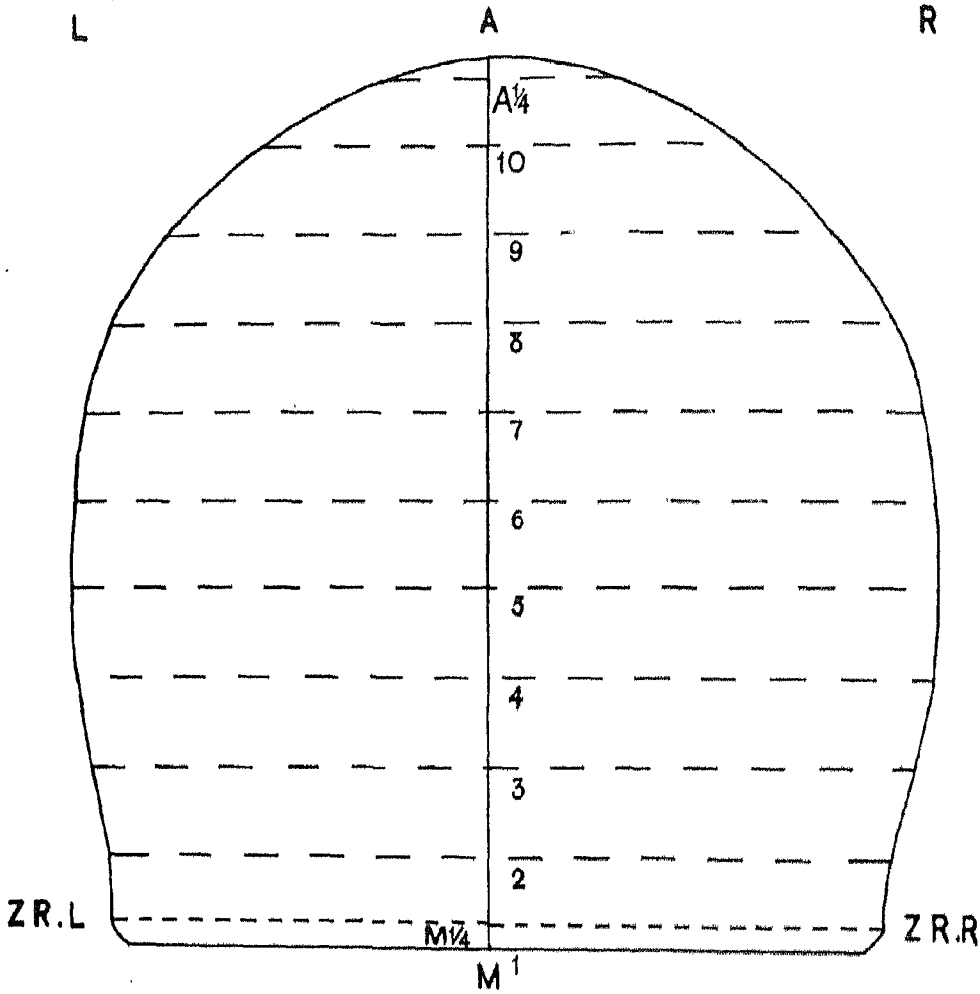


Fig. 5. Transverse type contour based on 89 female Lachish skulls.

by their mean measurements to the Upper Egyptian group, and the Lachish belongs to the Lower Egyptian (see Fig. 3). The superposed types show the greatest differences in calvarial breadth (transverse and horizontal sections), and decidedly smaller ones in calvarial lengths (horizontal and sagittal), as would have been anticipated. The Lachish facial skeleton is seen to be rather less prognathous than the Kerma or Badari, but the section of its nasal bones is the most projecting (cf. p. 145). The type for the predynastic series is the smallest in nearly all respects and particularly in facial height.

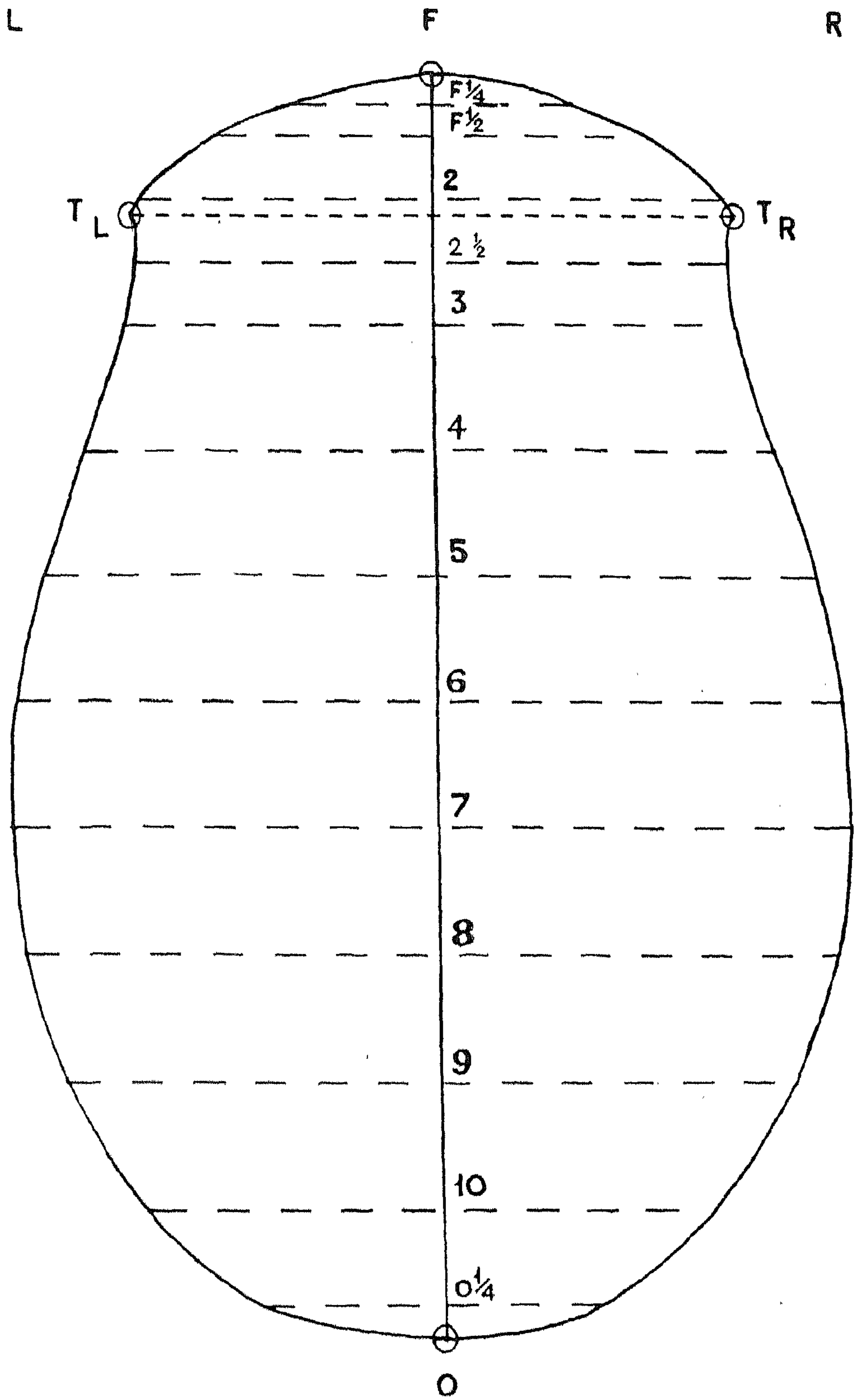


Fig. 6. Horizontal type contour based on 108 male Lachish skulls.

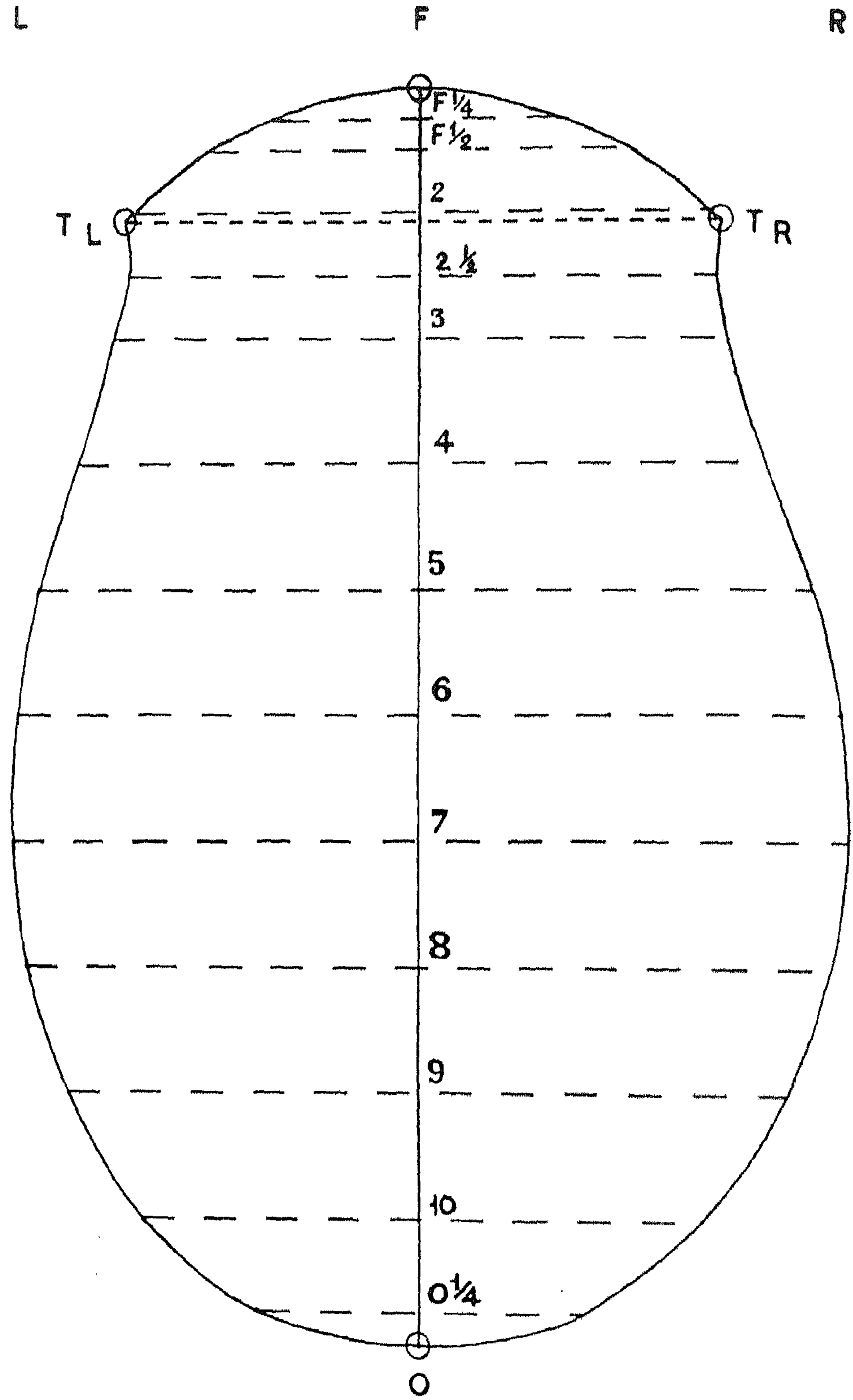


Fig. 7. Horizontal type contour based on 89 female Lachish skulls.

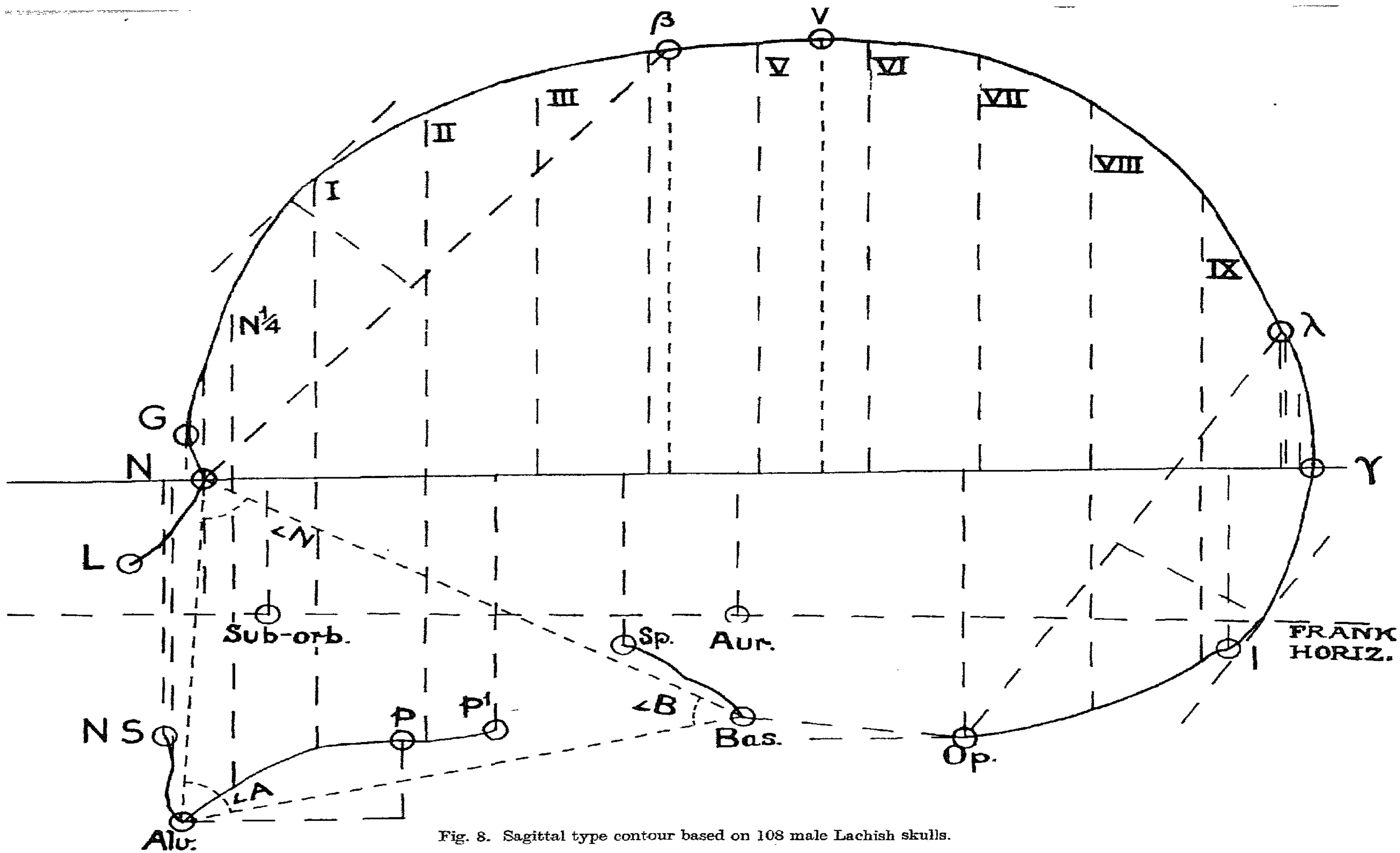


Fig. 8. Sagittal type contour based on 108 male Lachish skulls.

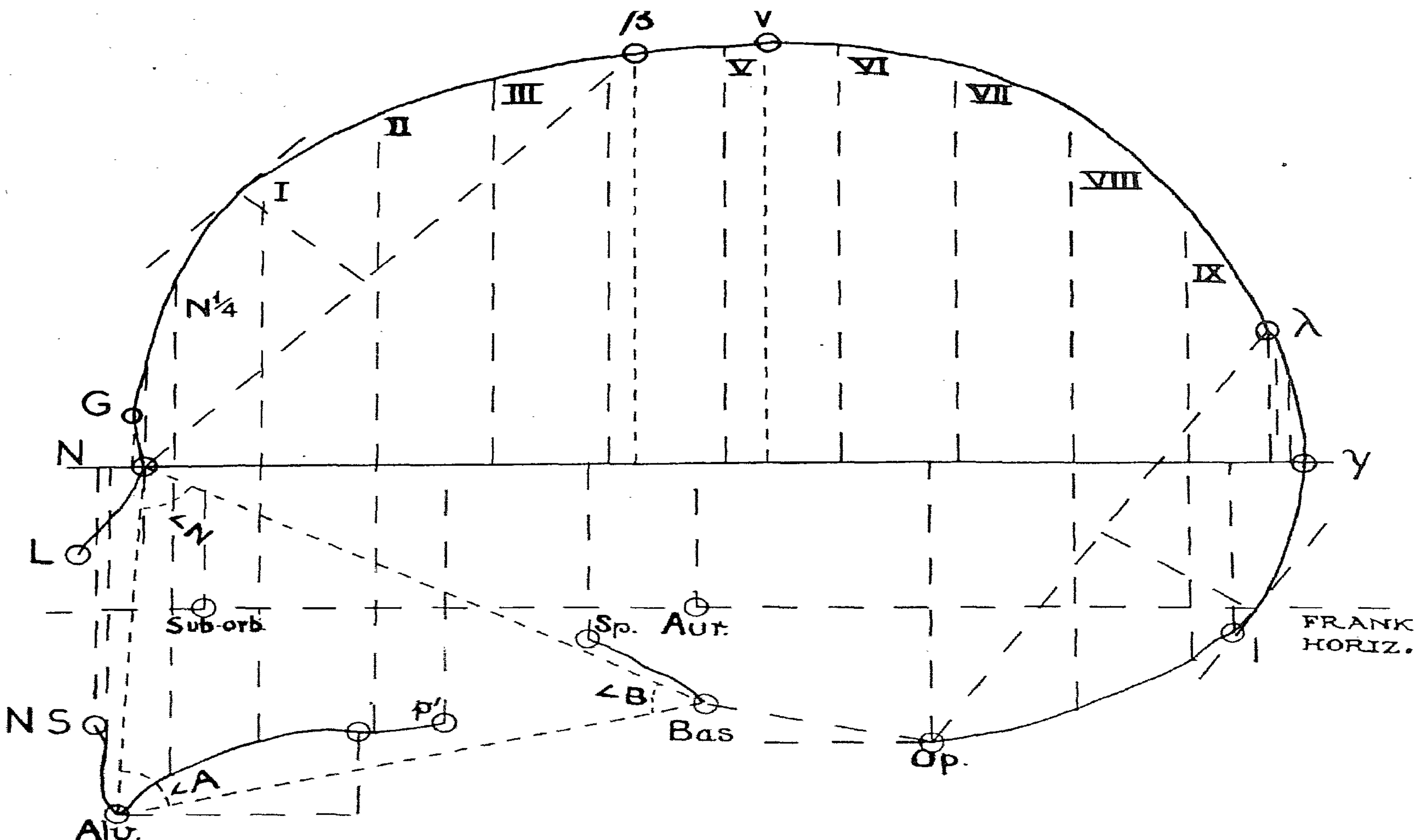


Fig. 9. Sagittal type contour based on 89 female Lachish skulls.

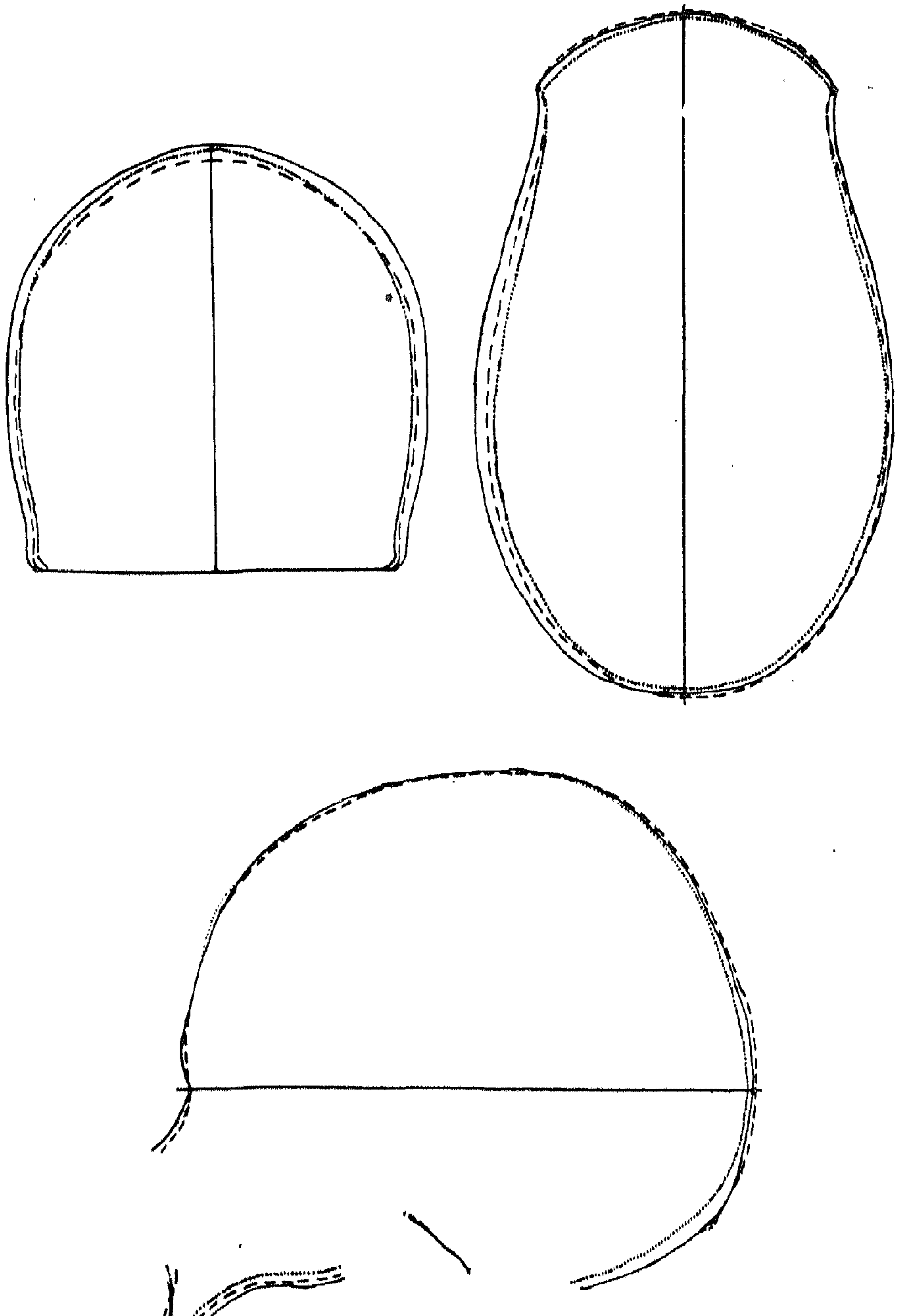


Fig. 10. Superposed male type contours of the Lachish (full line), Kerma (broken line) and Badari (dotted line) series. Scale half linear.

TABLE XV

Mean measurements of transverse contours of the Lachish skulls

Sex	*MA*	1*R*=1*L*	2*R*	2*L*	3*R*	3*L*	4*R*	4*L*	5*R*
♂	114·9 (108)	57·6 (108)	61·0 (108)	60·9 (108)	64·5 (108)	63·6 (108)	67·3 (108)	65·8 (108)	68·0 (108)
♀	109·1 (89)	55·4 (89)	59·0 (89)	58·4 (89)	62·3 (89)	61·1 (89)	65·3 (89)	63·3 (89)	65·8 (89)

Sex	5*L*	6*R*	6*L*	7*R*	7*L*	8*R*	8*L*	9*R*	9*L*
♂	66·5 (108)	67·9 (108)	66·0 (108)	66·3 (108)	64·3 (108)	62·0 (108)	60·1 (108)	53·1 (108)	51·4 (108)
♀	63·9 (89)	65·4 (89)	63·3 (89)	63·9 (89)	61·6 (89)	59·9 (89)	57·6 (89)	51·6 (89)	48·9 (89)

Sex	10*R*	10*L*	$A\frac{1}{4}R$	$A\frac{1}{4}L$	*ZR, R*		*ZR, L*	
					x	y	x	y
♂	38·4 (108)	36·5 (108)	18·5 (108)	17·0 (108)	3·1 (108)	60·2 (108)	3·1 (108)	60·7 (108)
♀	37·5 (89)	34·0 (89)	18·6 (89)	14·9 (89)	2·6 (89)	58·0 (89)	2·9 (89)	58·1 (89)

TABLE XVI

Mean measurements of horizontal contours of the Lachish skulls

Sex	FO	$F\frac{1}{4}R$	$F\frac{1}{4}L$	$F\frac{1}{2}R$	$F\frac{1}{2}L$	$2R$	$2L$	$2\frac{1}{2}R$	$2\frac{1}{2}L$	$3R$	$3L$
♂	184·3 (108)	23·0 (108)	22·6 (108)	35·0 (108)	35·4 (108)	47·6 (108)	47·9 (107)	48·5 (108)	48·5 (108)	49·6 (108)	50·1 (108)
♀	177·2 (89)	22·5 (89)	22·4 (89)	34·2 (89)	34·1 (89)	45·0 (89)	45·9 (89)	46·7 (89)	46·6 (89)	48·6 (89)	48·9 (89)

Sex	$4R$	$4L$	$5R$	$5L$	$6R$	$6L$	$7R$	$7L$	$8R$	$8L$
♂	55·8 (108)	56·8 (108)	62·9 (108)	63·6 (108)	66·9 (108)	68·0 (108)	68·0 (107)	69·2 (108)	65·4 (107)	67·1 (108)
♀	54·1 (89)	54·4 (89)	61·7 (88)	60·9 (89)	65·8 (88)	64·9 (89)	66·9 (89)	65·8 (89)	64·3 (89)	63·6 (89)

Sex	$9R$	$9L$	$10R$	$10L$	$O\frac{1}{4}R$	$O\frac{1}{4}L$	TR		TL	
							x	y	x	y
♂	58·1 (108)	60·4 (108)	44·4 (108)	47·7 (108)	25·1 (108)	29·4 (108)	20·9 (108)	49·3 (108)	20·9 (108)	49·3 (108)
♀	57·1 (89)	57·0 (89)	44·2 (89)	44·7 (89)	25·4 (89)	26·6 (89)	18·5 (89)	47·2 (89)	18·5 (89)	47·2 (89)

TABLE XVII

Mean measurements of sagittal contours of the Lachish skulls

Sex	$N\gamma$	Ordinates above $N\gamma$						
		$O=N$	$N\frac{1}{4}$	1	2	3	4	5
♂	182·3 (108)	21·6 (108)	37·6 (107)	59·4 (107)	72·5 (108)	80·1 (108)	84·2 (108)	85·7 (108)
♀	177·0 (89)	19·7 (89)	37·6 (89)	57·8 (89)	69·8 (89)	76·7 (89)	80·5 (89)	82·1 (89)

Sex	Ordinates above $N\gamma$						
	6	7	8	9	$\gamma\frac{1}{4}$	$\gamma\frac{1}{8}$	γ
♂	85·9 (107)	82·8 (107)	73·7 (108)	55·0 (108)	26·6 (108)	19·3 (108)	6·24 (108)
♀	82·2 (89)	78·6 (88)	69·9 (88)	50·6 (89)	24·4 (89)	17·1 (89)	5·0 (89)

Sex	Ordinates below $N\gamma$						Vertex	
	$O=N$	$N\frac{1}{4}$	1	2	8	9	x from N	y
♂	65·9 (87)	62·6 (95)	55·0 (105)	53·7 (97)	48·5 (106)	38·7 (108)	101·3 (108)	86·4 (108)
♀	62·1 (66)	58·5 (79)	52·2 (89)	51·2 (82)	47·3 (89)	37·7 (89)	94·5 (89)	82·9 (89)

Sex	Bregma		Glabella		λ		Suborb.	
	x from N	y	x from N	y	x from γ	y	x from N	y
♂	76·2 (108)	84·3 (108)	2·7 (108)	8·9 (108)	5·1 (108)	27·1 (108)	10·2 (108)	27·8 (108)
♀	74·7 (89)	81·0 (89)	1·4 (89)	9·4 (89)	5·3 (89)	25·4 (89)	9·6 (89)	27·2 (89)

TABLE XVII (*cont.*)

Mean measurements of sagittal contours of the Lachish skulls (cont.)

Sex	Aur. Pt.		Opisthion		Inion		Basion	
	x from N	y	x from γ	y	x from γ	y	From γ	From N
♂	87·7 (108)	29·3 (108)	57·3 (97)	54·3 (97)	13·8 (107)	36·5 (107)	106·3 (96)	101·8 (96)
♀	84·0 (89)	27·8 (89)	57·3 (79)	53·5 (79)	11·4 (89)	32·8 (89)	102·6 (74)	97·1 (74)

Sex	Alv. Pt.		Nose				
	From Bas.	From N	(i)	(ii)	(iii)	$\angle \gamma NL$	NL
♂	94·2 (62)	69·7 (66)	1·3 (100)	4·7 (70)	7·0 (33)	125°·3 (13)	21·0 (13)
♀	91·1 (38)	66·2 (44)	1·2 (83)	3·5 (67)	7·1 (29)	122°·8 (11)	19·4 (11)

Sex	Palate				Frontal		Occipital	
	P'		P		Max. Sub. to $N\beta$		Max. Sub. to λ Op.	
	x from N	y	x from Alv.	y	x from N	y	x from λ	y
♂	47·9 (76)	51·6 (76)	36·1 (65)	16·0 (65)	50·9 (107)	26·6 (107)	50·1 (97)	28·6 (97)
♀	45·7 (56)	49·6 (56)	36·1 (44)	15·0 (44)	49·0 (89)	26·7 (89)	46·8 (79)	27·5 (79)

Sex	Sp.		NS		Vert. Tan. below NS		Sub. from ½ Bas.-Sp. chord above
	x from N	y	x from N	y	x from N	y	y
♂	68·9 (105)	34·9 (105)	6·3 (36)	52·0 (36)	5·3 (67)	56·9 (67)	0·4 (96)
♀	67·7 (83)	33·8 (83)	6·9 (25)	49·4 (25)	5·0 (51)	53·9 (51)	0·4 (74)

9. Comparisons of the Lachish series with series of crania and living people other than Egyptian

The comparisons made above have shown that the Lachish series is very similar in type to several Ancient Egyptian series. In fact it bears as close a resemblance to some of these as they do in general to one another. Hence it may be concluded that the population of Lachish in the year 700 B.C. was primarily, at least, of Egyptian origin. The cranial evidence from other countries of the Near East is very meagre. Comparisons are made with the only other Palestinian series of any length, and with a few series of living people, in this section.

A considerable number of human skeletons was excavated by Prof. R. A. S. Macalister at Gezer, 17 miles south-east of Jaffa, from 1902 to 1905 and 1907 to 1909. These are dealt with in a chapter in his report (1912). Material of the pre-Semitic period was very fragmentary, and no measurements of it could be taken. Five indices for a series of skulls of the Semitic periods are given in the form of frequency distributions only. Division is made between a series for the first and second periods together, and a series for the third and fourth periods together, in the case of the cephalic index. A distribution of reconstructed statures is also provided. The material was apparently unsexed, and it is not clear how the absolute frequencies could be determined from the diagrams. All that can be said is that the average cephalic index for both series of unsexed skulls is about 76. The Lachish male mean is 74·3, and the female 75·5.

There appear to be no published records for any other long series of skulls of any date from Palestine, or any neighbouring country except Egypt. Measurements of small numbers of ancient Jewish and Phoenician and modern Arab specimens have been given, but these are practically worthless for statistical purposes. There are no adequate data for Ancient Jewish skulls from any locality. The longest modern series representing this people is one published by Prof. J. Matiegka (1926) of seventeenth-century Jews buried in Prague. The average cephalic index for the fifty-three male skulls is 82·0, which is sufficient to show that there can be no close connexion with the Lachish people (74·3), or with any of the Ancient Egyptian groups, for which the highest index is 76·0. The coefficient of racial likeness was computed for twenty-two characters between the male Jewish series from Prague ($\bar{n} = 32{\cdot}8$) and the Lachish series ($\bar{n} = 202{\cdot}8$): a reduced value of 52·6 is found. This is greater than the maximum found between any pair of the series of Egyptian type, including the Lachish (see Table XII). It is not suggested, of course, that the Palestinian Jews in 700 B.C. were necessarily of the same type as Jews in Prague in the seventeenth century.

Measurements of a certain number of Jewish people in Palestine have been published, and only the cephalic index for series of men will be considered here. No accurate comparisons with cranial data can be given in the case of any

other characters recorded. Weissenberg (1909) gives a mean value of 79·8 for fourteen men in Galilee. All the remaining series of Jews in Palestine represent the Samaritans, the sources and mean indices being:

Weissenberg (1909)—76·2 (20); Kappers (1934)—77·2 (27);
Szpidbaum (1927)—77·6 (27); Huxley (1906)—78·1 (35);
Genna (1938)—79·1 (39).

In view of the small sizes of the series, these averages accord remarkably well. The pooled mean for the 148 men is 77·9.

Weissenberg (1909) has also given cephalic indices of 76·9 for twenty-five fellahin measured near Jaffa, and 75·7 for thirty in the Safed district, or a pooled mean of 76·2 for the fifty-five men. This value differs markedly from that of 81·6 given by Kappers (1934) for 139 Arabs from "towns north of the line Jaffa–Jericho". There appear to be some well-marked regional differences between Arabs in different localities in Palestine.

The mean cephalic index for the male Lachish series of skulls is 74·3, which can be supposed to correspond to a value of about 76·3 in the living. The latter is decidedly less than the mean for one series of Arabs, rather less than the Jewish means, and practically identical with that for the Arabs measured by Weissenberg. It is quite unsafe to lay stress on comparisons of a single character, but those made show that a population with the same cephalic index as that of the ancient inhabitants of Lachish is living in Palestine to-day. The possibility that the pre-Christian type has persisted until modern times is not precluded, though far more evidence—and particularly that of later series of skulls from the country—will be needed to disclose its racial history in any detail.

10. MANDIBLES AND LONG BONES OF THE LACHISH SERIES

In all there are seventy-six mandibles in the Lachish series—fifty-six from Tomb 120 and the remainder from the three other tombs—most of them being defective to some extent. Of the total, one is associated with an adult male cranium, and nine with adult female crania. The remaining sixty-six are unassociated, and they were sexed by anatomical appreciation. Remarks on a few of the specimens noted as being anomalous are in § 4 above. Measurements were taken of the adult bones in accordance with the biometric technique (Morant, Collett & Adyanthāya, 1936). There are thirty-four male and thirty-five female specimens. Means for these are given in Table XVIII, and comparisons with other material would not be profitable, as it has been shown (Cleaver, 1937) that considerably larger numbers would be required to reveal small racial differences in type. Neither the measurements nor the appearance of the Lachish mandibles suggest any clear divergence from Ancient Egyptian types.

TABLE XVIII

Mean measurements of Lachish mandibles

Character	Male	Female	Character	Male	Female
w_1	118·6 (7)	114·9 (7)	c_rh	64·3 (11)	60·0 (16)
c_yl	20·5 (17)	19·4 (18)	m_2h	26·9 (14)	24·5 (14)
rb'	32·6 (26)	29·9 (28)	$M\angle$	125°·9 (14)	124°·5 (15)
m_2p_1	27·4 (26)	27·2 (23)	$R\angle$	71°·6 (5)	72°·1 (9)
h_1	34·8 (20)	31·8 (19)	$C\angle$	74°·6 (11)	72°·7 (13)
zz	44·8 (31)	43·2 (33)			
c_pl	73·1 (14)	70·1 (15)	$100\ c_rh/ml$	60·7 (5)	60·5 (6)
rl	59·75 (14)	54·9 (15)	$100\ c_rc_r/ml$	86·6 (4)	90·5 (5)
g_0g_0	97·1 (11)	85·5 (11)	$100\ g_0g_0/c_pl$	133·3 (11)	122·9 (11)
c_rc_r	91·1 (4)	92·1 (8)	$100\ rb'/rl$	54·7 (14)	53·9 (14)
ml	104·8 (8)	100·3 (8)	$100\ g_0g_0/c_rc_r$	101·4 (4)	92·1 (6)

Other parts of the skeleton are only represented in the Lachish series (all tombs) by two sacra and nearly 200 long bones, many of which are incomplete. These are not associated together and no attempt was made to sex them. The maximum lengths of the adult long bones were determined, and means for them are given in Table XIX. As far as can be seen from these constants, the Lachish people were rather short, but no approximation of any value to the average statures of the men and women can be given. One ulna (No. 2) has a healed fracture of the lower shaft, and one femur (No. 39) has condyles affected by arthritis.

TABLE XIX

Means of the maximum lengths of unsexed adult long bones of the Lachish series

	Femora (oblique)	Tibiae (oblique)	Humeri (oblique)	Radii	Ulnae	Clavicles
R	427·9 (18)	347·7 (16)	298·5 (11)	243·7 (6)	264·6 (5)	147·25 (4)
L	430·45 (20)	372·25 (8)	300·1 (15)	239·8 (5)	266·0 (6)	154·0 (1)

11. Summary and conclusions

The skeletal remains reported on in this paper for the Trustees of the late Sir Henry Wellcome were collected at Tell Duweir (Lachish), twenty-five miles south-west of Jerusalem, by the Wellcome-Marston Expedition to the Near East from 1933 to 1936. The bones were found in four adjoining tomb chambers and they are assigned to the seventh and eighth centuries B.C. In all there are 695 crania, the majority of which are more or less imperfect, and much smaller numbers of mandibles and other bones of the skeleton (see table on p. 103). Of the crania 360 were judged to be adult male and 274 adult female, the

remaining sixty-one being immature. The origin of the collection is discussed, and it is concluded that the remains are probably those of people who died as the result of some catastrophe. The frequencies of occurrence of different states of closure of the principal calvarial sutures show, from comparisons with other cranial series, that the adults from Lachish were younger, on the average, than cemetery populations are expected to be. Very few aged individuals were interred in the tombs. The normal order of closing of the sutures was sagittal—coronal—lambdoid. Remarks on unusual conditions and anomalies are given, the most interesting specimens being three trepanned skulls; two showing marked artificial deformation and six suspected to have been deformed artificially; a series of seventeen showing premature closing of the sagittal suture without clear deformation except in three cases; one believed to be distorted owing to premature closing of the coronal suture; one with absence of the right auricular passage; and one with an extensive diseased area on the vault.

Statistics regarding the loss of teeth before death show that they were remarkably well preserved. Remarks on dental anomalies are given. The most interesting skull from this point of view is one which was found to have a metal filling in one of its molars, presumably acquired by accident.

Judging from comparisons of the measurements, there is no reason to doubt that the series from the four tombs represent precisely the same population, and the differences found between the male and female adult and juvenile constants are no greater than those expected in such a case. The variabilities and sex ratios of the total series (combining skulls from all tombs) are quite unexceptional.

Comparisons are made between the Lachish and twenty-one Ancient Egyptian and allied series of skulls by the method of the coefficient of racial likeness, and a classification of the material is presented. The relationships found suggest that the population of the town in 700 B.C. was entirely, or almost entirely, of Egyptian origin, very close connexions with some contemporary Egyptian series being found. They show, further, that the population of Lachish was probably derived principally from Upper Egypt. Comparisons of measurements considered singly indicate that the Lachish cranial type has no features which would be unusual for an Ancient Egyptian type, other than the prominence of its nasal bones and the curvature of its malar bones. Transverse, horizontal, and sagittal type contours based on 108 male and eighty-nine female Lachish skulls are given, and it is shown that they are very similar to some previously provided for Ancient Egyptian series. There are no good records for any series of skulls from Palestine other than the Lachish. Its mean cephalic index accords fairly well with that given for a series of living Palestine Arabs, and it is close to that for Samaritans. The possibility that the Lachish people have persisted until to-day is not precluded, but far more evidence would be required to substantiate such a hypothesis. The series of mandible and long bones from Lachish are too small to be of any value for comparative purposes. As far as can be seen the people were rather short.

APPENDICES

I. *Definitions of skull measurements taken*

Measurements of the Lachish material were taken in accordance with biometric practice. The definitions of cranial points given by Buxton & Morant (1933), and those of mandibular measurements given by Morant, Collett & Adyanthāya (1936), were followed. The contractions below are used to denote measurements in the tables and text, and their numbers in Martin's list are given.

C = capacity in c.c. It was not possible to determine the capacities of any of the Lachish skulls directly, owing to the fact that the interiors of the brain-boxes are coated with mud and wax which cannot be removed. The reconstruction formulae using L, B, and H' given by Pearson & Stoessiger (1927) were applied to give the estimates in Table VI, and these were not used in computing coefficients of racial likeness. L = maximum glabella-occipital length (M. 1). B = maximum horizontal breadth (M. 8). H' = basio-bregmatic height (M. 17). LB = basion to nasion (M. 5). B' = minimum frontal breadth (M. 9). S = arc nasion to opisthion (M. 25). S_1 = arc nasion to bregma (M. 26). S_2 = arc bregma to lambda (M. 27). S_3 = arc lambda to opisthion (M. 28). S'_1 = chord nasion to bregma (M. 29). S'_2 = chord bregma to lambda (M. 30). S'_3 = chord lambda to opisthion (M. 31). U = horizontal circumference measured through the ophyron and directly above the superciliary ridges (M. 23*a*). $\beta Q'$ = transverse circumference from one auricular point to the other, passing through bregma (M. 24). *fml* = basion to opisthion (M. 7). *fmb* = maximum breadth of foramen magnum (M. 16). $G'H$ = nasion to alveolar point (M. 48). GL = basion to alveolar point. GB = facial breadth between lowest points on zygomatic-maxillary sutures (M. 46). J = maximum breadth between zygomatic arches (M. 45). NH, L = nasion to lowest edge of pyriform aperture on the left side. NB = maximum breadth of pyriform aperture (M. 54). O_1L = maximum breadth of left orbit (M. 51). O_2L = maximum height of left orbit (M. 52). G'_1 = length of palate from orale to staphylion (M. 62). G_2 = breadth of palate between inner alveolar walls of second molars (M. 63). OH = maximum projection from biporial axis in the transverse vertical plane, measured on transverse contour. SC = simotic chord, minimum breadth of nasal bones (M. 57). SS = subtense of simotic chord. Measurements of the left malar bones, taken in accordance with Woo's instructions (1937), are; Ml_1 = minimum horizontal arc. Ml_2 = minimum vertical arc. $C(ml)$ = chord between terminals of horizontal arc. $S(ml)$ = maximum subtense from the chord. The occipital index is the only one which needs definition: it is

$$Oc.I. = 100\frac{S_3}{S'_3}\sqrt{\left(\frac{S_3}{24(S_3 - S'_3)}\right)}.$$

Values for the individual skulls were found with the aid of Miss Tildesley's table of the function (*Biometrika*, **13** (1921), 261–2). $P\angle$ = profile angle between Frankfurt horizontal plane and the chord joining nasion to alveolar point. $N\angle$, $A\angle$ and $B\angle$ are the angles of the triangle of which the nasion, alveolar point and basion are the apices. w_1 = maximum breadth outside condyles. c_yl = maximum length of the left condyle. rb' = minimum antero-posterior "breadth" of the left ramus. m_2p_1 = chord between the points on the outer left alveolar margin from the middle of the second molar to the middle of the first premolar. h_1 = symphyseal height from intradental to the point farthest removed from it in the symphyseal plane. zz = minimum chord between the anterior margins of the right and left *foramina mentalia*. c_rc_r = coronial breadth from right coronion to left coronion. $M\angle$ = mandibular angle. c_pl = projective length of the corpus. rl = projective length of the left ramus. g_0g_0 = chord from left gonion to right gonion. ml = maximum projective length of the mandible. c_rh = projective height of the left coronoid process. m_2h = projective height of the corpus at the middle point of the outer alveolar margin of the second left molar. $R\angle$ = angle of condylar-coronoidal line with ramus tangent. $C'\angle$ = angle between the standard horizontal plane and the line joining the infradental to the most anterior point in the standard sagittal plane of the symphysis.

ipteric bone R; 5 molars lost

L obliterated; palate defective

ore death; fused nasal bones;

probably no teeth lost before

death; M3's absent; abscess

R; 1 molar lost before death.

mporal squama and parietal;

ore death, unusual absorption

ost teeth lost before death
healed wound R parietal
molars lost before death
absent
ve
molar lost before death, M3's

132	183	133	96·9	138·5
133	186·5	141	94·0	134·5
134	179·5	133·5	96·2	129·5
135	191·5	135	97·1	132
136	182·5	145	97·3	140·5
137	179	139	95·2?	137
138	177	129	89·2	128
139	187·5	128	92·7	—
140	190·5	132	94·8	134
141	181	126	90·1	129·5
142	173·5	131·5	89·9	131
143	197·5	138	95·0	143

rks

mastoid obliterated; parieto
icle embedded in palate
emporal articulation L
lars lost before death
suture between ex- and sup
beginning to close a-occipital R

nissing
nissing
nearly obliterated, sagittal
sent
olars lost before death; fused
closing, sagittal beginning rietal bones R
sent al R
; large wormian bone at ast

lefective; epipteric bone L; pra-occipitals R and L; wormian bone be-
alf of palate missing
f horizontal suture of inter bones between temporal squama and parie-
and unerupted, obstructing
al cusp on M 3, R
pipteric bone L, and 2 R
nissing
and sagittal closed; additi

; front teeth crowded; M 3
dult; M 3's not complete
t asteria; 2 epipteric bone
-mastoid L obliterated; p
and sagittal closed; fused al R
; palate missing es R; L side of palate missing
ted palate; 1 molar lost b
isor and 4 molars lost befo
lar lost before death; M 3

pipteric bone R and L; p
; trace of sutures betwee ing
beginning to close; 4 mo
 ital R and L
lefective
pipteric bones R and L;

dult; single epipteric bo
lar lost before death
l bone missing; 1 prem teric bone L; single epipteric bone R; basal
beginning to close
in K, R and L arietal bones R
lar lost before death

pipteric bone R; palate tubercles of Carabelli
closing, sagittal closed nes L; healed injury above L orbit
beginning to close; normally large and L abnormally small
ive
all pre-condyles
lar lost before death, d-
bones. Mandible: no
r margin, malocclusion f open
isors lost before death obliquely behind I 1, R
defective, nearly all tee-

defective
parietal R
defective
 skeletons
, canine, or premolar
l to R

LIST OF PLATES

Plate I. Views of the interior of Tomb 120.
- A. Entrance of tomb showing skulls round side.
- B. Interior of tomb with skulls collected round side.

Plate II. Tomb 120, collections of skulls *in situ*.
- A. Skulls collected round wall.
- B. A closer view of skulls seen to the left of the beam in Plate IA.

Plate III. Exceptional crania.
- A. A female skull (No. 485) showing extensive burnt patch.
- B. A male skull (No. 108) showing injuries probably inflicted shortly before death: see p. 104.
- C, D. Male skulls (No. 156, Tomb, 107 and No. 179, Tomb 116) of exceptional type; no other specimens of this type were found.

Plate IV. A male cranium (No. 340) with hole in right parietal, probably a trepan, and a sword-cut on the same bone close to the coronal suture.

Plate V. Male crania with trepanned openings.
- A. No. 115. There is a septic area round the trepan in the right parietal. A cut probably made by the trepanner can be seen by the lateral outline of the left parietal.
- B. No. 114.

Plate VI. Crania showing marked artificial deformation.
- A. A male cranium (No. 381).
- B. A female cranium (No. 673).

Plate VII. *Norma lateralis* views of five male crania (A–E) believed to be artificially deformed to a slight extent, and of a male cranium (F), probably deformed owing to premature closing of the coronal suture. A, No. 378; B, No. 375; C, No. 377; D, No. 376; E, No. 379; F, No. 380.

Plate VIII. *Norma verticalis* views of three anomalous crania.
- A. A male cranium (No. 380) with coronal suture obliterated. A *norma lateralis* view of this deformed specimen is shown in Plate VIIF.
- B. A female cranium (No. 454) with healed injury on left side of the frontal bone.
- C. A metopic female cranium (No. 670) with wound on left parietal bone and the sagittal suture prematurely obliterated.

Plate IX. *Norma lateralis* views of a typical female (No. 388) and a typical male (No. 28) cranium of the Lachish series. Other views of these two specimens are shown in Plates X–XII. They were selected from among the more complete skulls on account of the fact that their measurements of shape show no marked divergences from the means for the series. In the case of No. 388, none of these measurements differ from the female mean by more than the standard deviation of the distribution. The same is true for No. 28 in comparison with the male means. All the photographs of the typical skulls are approximately 0·6 natural size (linear dimensions).

Plate X. *Norma facialis* views of a typical female (No. 388) and a typical male (No. 28) cranium of the Lachish series.

Plate XI. *Norma verticalis* views of a typical female (No. 388) and a typical male (No. 28) cranium of the Lachish series.

Plate XII. *Norma occipitalis* views of a typical female (No. 388) and a typical male (No. 28) cranium of the Lachish series.

Plate XIII. Crania showing premature obliteration of the sagittal suture.

A. A female cranium (No. 672) showing complete obliteration of the suture with distortion.

B. A male cranium (No. 364) showing complete obliteration of the suture with distortion.

C. A female cranium (No. 667) showing complete obliteration of the suture with distortion.

D. A male cranium (No. 366) showing complete obliteration of the suture without apparent distortion.

E. A female cranium (No. 577) showing almost complete obliteration of the suture and post-coronal constriction.

Plate XIV. Three male crania with sutural anomalies.

A. No. 299. Large wormian bone in right side of coronal suture.

B. No. 60. Two symmetrical interparietal bones of unusual form.

C. No. 330. Left side, showing temporal bone largely fused to parietal. The right side of this specimen is affected in a similar way.

Plate XV. Anomalous regions of two male crania.

A. No. 324. Complete absence of right auricular passage: twice natural size. The left auricular passage is normal.

B. No. 301. Complete absence of left jugular foramen: 1·3 times natural size. The right foramen is normal.

Plate XVI. A male cranium (No. 382) of unusual form, possibly affected by hydrocephaly. A, *Norma lateralis*; B, *Norma facialis*; C, *Norma verticalis*.

Plate XVII. A female cranium (No. 662) with large diseased area on frontal bone (see p. 115). A, *Norma lateralis*; B, *Norma facialis*; C, *Norma verticalis*.

Plate XVIII. Exceptional crania.

A. A male cranium (No. 5) with large healed wound on frontal bone.

B. A male cranium (No. 1) with diseased area on right parietal.

C. A metopic female cranium (No. 419) with injury to the right frontal bone.

D. A female cranium (No. 513) with depression in left parietal.

Plate XIX. A female cranium (No. 518) showing a tooth (upper right second molar) with a filling presumed to be adventitious.

A. The palate (1·2 diameters).

B. The second molar with the filling *in situ* (6·0 diameters).

C. A skiagram (2·7 diameters) showing the depth of the filling.

D. The three molars (2·7 diameters) after removal of the filling in the second molar.

The skiagram in this plate and others in Plates XXII and XXIII were kindly provided by Mr C. Bowdler Henry, M.R.C.S.

Plate XX. A female skull (No. 437) with anomalous jaws.

A. The mandible from above, showing denticles outside the dental arch.

B. The palate showing a diastema between the central incisors, and diastemae between the lateral incisors and canines.

C. The right side of the mandible showing denticles.

Plate XXI. Palates with anomalous dentitions.

A. A female cranium (No. 445) with sockets for three incisors only.'

B. A male cranium (No. 132) with diastemae between canines and premolars, and third molars absent.

C. A juvenile cranium (No. 705) with supernumerary tooth behind the right central incisor.

D. A female cranium (No. 401) with grossly deflected left canine.

Plate XXII. Photographs and skiagrams of jaws with anomalous dentitions.

A. A female cranium (No. 506) showing incomplete eruption of the second and third right molars.

B. Skiagram of the same specimen (A) showing that the roots of the partially erupted teeth were fully formed.

C. A female cranium (No. 383) showing a denticle in the palate.

D. Skiagram of the same (C) showing the limits of the denticle and its crypt.

E. A female cranium (No. 496) with impaction of the upper left third molar.

F. Skiagram of the right side of a mandible showing two denticles in the corpus, female (No. 1056).

Plate XXIII. Jaws with anomalous dentitions or cysts.

A. A skiagram of a juvenile jaw (No. 1068), right side, showing canine and first premolar unerupted.

B. A skiagram of the same mandible (A), left side, showing the same teeth on this side unerupted (see p. 120).

C. A male skull (No. 72) showing the third molar on the left side in abnormal position.

D. Occlusal view of the same jaw (C) showing the abnormal position of the third molar, rotation of the second premolar on the left side, a retained milk canine, and absence of the third molar on the right side (see description on p. 119).

E. A female cranium (No. 469) showing a large cyst in the right molar region.

F. A female cranium (No. 467) showing a large cyst in the anterior part of the palate penetrating to the nasal aperture.

REFERENCES

BATRAWI, A. M. EL (1935). "Report on the human remains." *Miss. Archéol. de Nubie, 1929–34.* Cairo.

BONIN, G. VON & MORANT, G. M. (1938). "Indian races in the United States. A survey of previously published cranial measurements." *Biometrika,* **30**, 94–129.

BUXTON, L. H. D. & MORANT, G. M. (1933). "The essential craniological technique. Part I. Definitions of points and planes." *J. Roy. Anthrop. Inst.* **63**, 19–47.

CLEAVER, F. H. (1937). "A contribution to the biometric study of the human mandible." *Biometrika,* **29**, 80–112.

COLLETT, M. (1933). "A study of the 12th and 13th dynasty skulls from Kerma." *Biometrika,* **25**, 254–84.

DAVIN, A. G. & PEARSON, K. (1924). "On the biometric constants of the human skull." *Biometrika,* **16**, 328–63.

DINGWALL, E. J. (1931). *Artificial Cranial Deformation.* London.

DOUBLE, A.-F. LE (1903). *Traité des Variations des Os du Crâne de l'Homme.* Paris.

GENNA, G. E. (1938). *Spedizioni Scientifiche del Comitato Italiano per lo Studio dei Problemi della Popolazione. Seconda Spedizione: i Samaritani. Anthropologia,* **1**. Rome.

HOOKE, B. G. E. (1926). "A third study of the English skull with special reference to the Farringdon Street crania." *Biometrika,* **18**, 1–54.

HRDLIČKA, A. (1932–3). "Seven prehistoric American skulls with complete absence of external auditory meatus." *Amer. J. Phys. Anthrop.* **17**, 355–77.

HUXLEY, H. M. (1906). "Zur Anthropologie der Samaritaner." *Z. Demogr. Statist. Jud.* **2**, 137–9.

KAPPERS, C. U. ARIENS (1934). *An Introduction to the Anthropology of the Near East in Ancient and Recent Times.* Amsterdam.

MACALISTER, R. A. S. (1912). *Excavations at Gezer,* **1**. London.

MACDONELL, W. R. (1904). "A study of the variation and correlation of the human skull, with special reference to English crania." *Biometrika,* **3**, 191–244.

MARTIN, E. S. (1936). "A study of an Egyptian series of mandibles with special reference to mathematical methods of sexing." *Biometrika*, **28**, 149–78.

MATIEGKA, J. (1926). "On the craniology of the Jews. I. The skulls from the Old Cemetery, Prague, V." *Anthrop., Prague*, **4**, 163–219.

MORANT, G. M. (1925). "A study of Egyptian craniology from prehistoric to Roman times." *Biometrika*, **17**, 1–52.

—— (1928). "A preliminary classification of European races based on cranial measurements." *Biometrika*, **20B**, 301–75.

—— (1935). "A study of predynastic Egyptian skulls from Badari based on measurements taken by Miss B. N. Stoessiger and Prof. D. E. Derry." *Biometrika*, **27**, 293–308.

MORANT, G. M., COLLETT, M. & ADYANTHĀYA, N. K. (1936). "A biometric study of the human mandible." *Biometrika*, **28**, 84–122.

MORANT, G. M. & HOADLEY, M. F. (1931). "A study of the recently excavated Spitalfields crania." *Biometrika*, **23**, 191–248.

PARRY, T. WILSON (1936). "Three skulls from Palestine showing two types of primitive surgical holing." *Man*, **36**, 170–1.

PEARSON, K. & STOESSIGER, B. N. (1927). "On further formulae for the reconstruction of cranial capacity from external measurements of the skull." *Biometrika*, **19**, 211–14.

STARKEY, J. L. (1936). "Discovery of skulls with surgical holing at Tell Duweir, Palestine." *Man*, **36**, 169–70.

STOESSIGER, B. N. (1927). "A study of the Badarian crania recently excavated by the British School of Archaeology in Egypt." *Biometrika*, **19**, 110–50.

STOESSIGER, B. N. & MORANT, G. M. (1932). "A study of the crania in the vaulted ambulatory of St Leonard's Church, Hythe." *Biometrika*, **24**, 135–202.

SZPIDBAUM, H. (1927). "Die Samaritaner." *Mitt. anthrop. Ges. Wien*, **57**, 139–58.

THOMSON, A. & MACIVER, D. RANDALL (1905). *The Ancient Races of the Thebaid*. Oxford.

WEISSENBERG, S. (1909). "Die autochtone Bevölkerung Palastinas in anthropologischer Beziehung." *Z. Demogr. Statist. Jud.* **5**, 129–39.

WOO, T. L. (1930). "A study of seventy-one ninth dynasty Egyptian skulls from Sedment." *Biometrika*, **22**, 65–93.

—— (1937). "A biometric study of the human malar bone." *Biometrika*, **29**, 113–23.

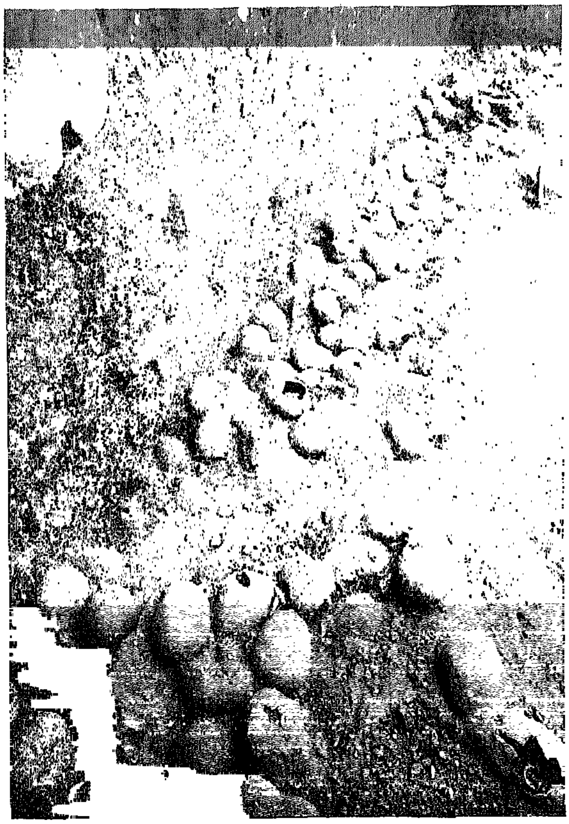
WOO, T. L. & MORANT, G. M. (1932). "A preliminary classification of Asiatic races based on cranial measurements." *Biometrika*, **24**, 108–34.

—— —— (1934). "A biometric study of the 'flatness' of the facial skeleton in man." *Biometrika*, **26**, 196–250.

A

B

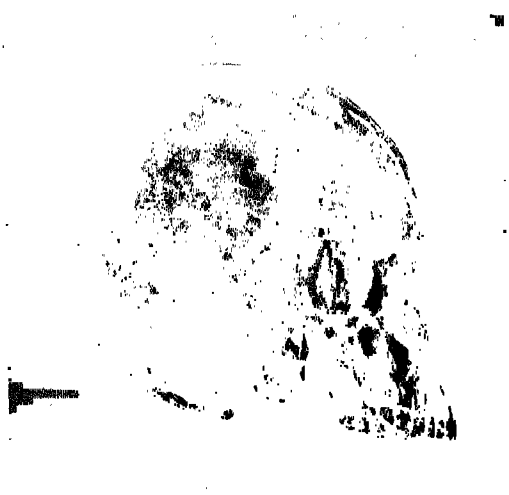
Views of the interior of Tomb 120.



A

B

Tomb 120, collections of skulls *in situ*.



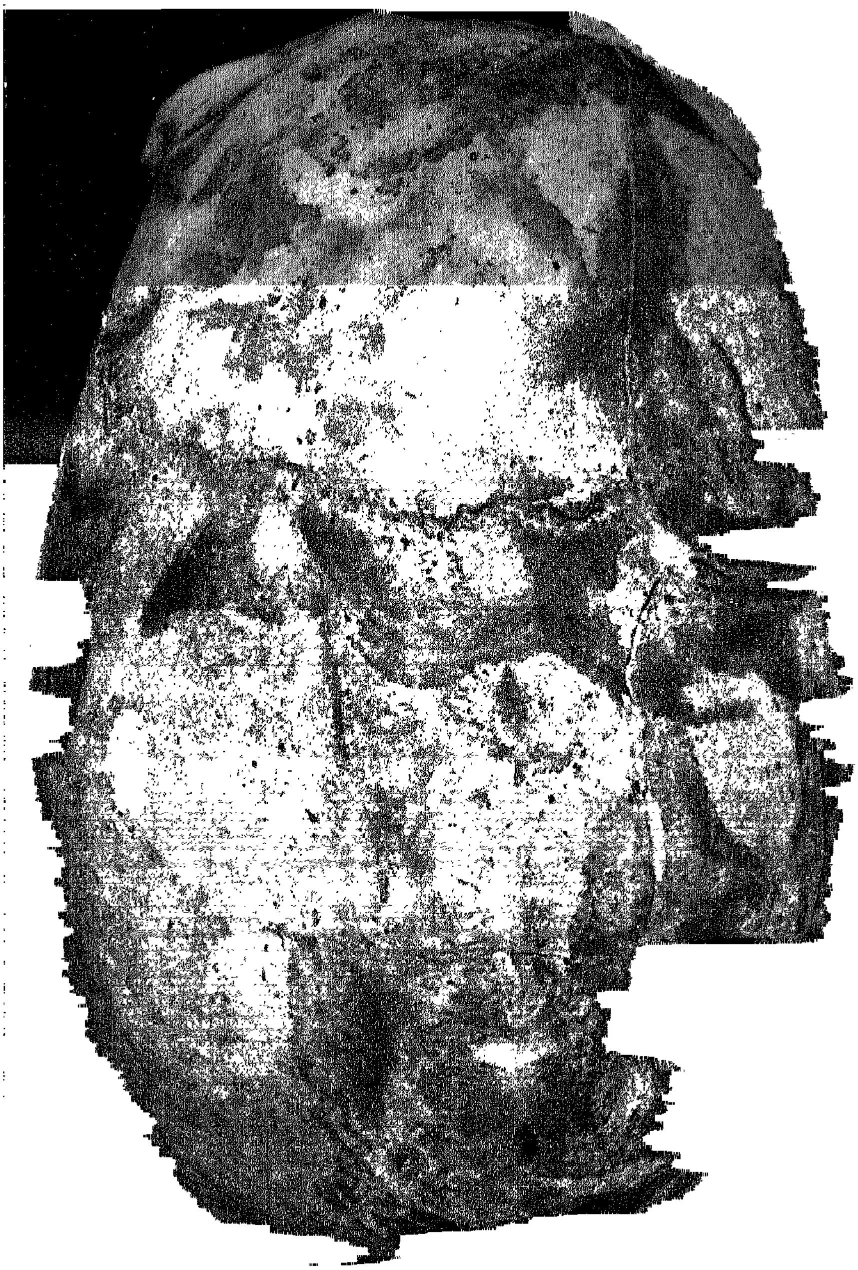
A. A female skull (No. 485) showing extensive burnt patch.

B. A male skull (No. 108) showing injuries which probably caused death.

C. A male skull (No. 156) of exceptional type.

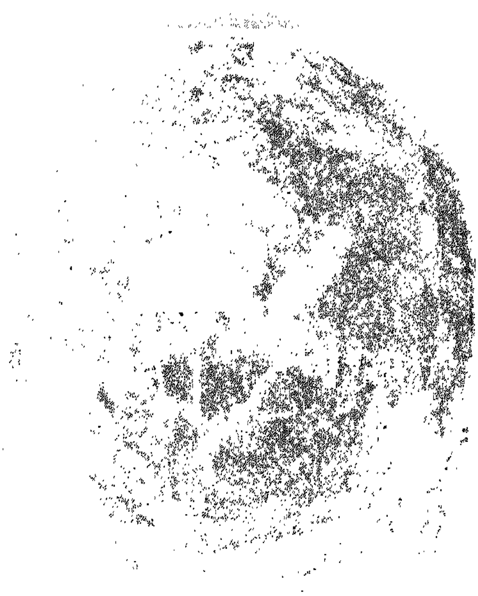
D. A male skull (No. 179) of exceptional type and similar to C.

Exceptional crania.



A male cranium (No. 340) with hole in right parietal, probably a trepan, and sword-cut on same bone close to coronal suture.

A. No. 115.



B. No. 114.

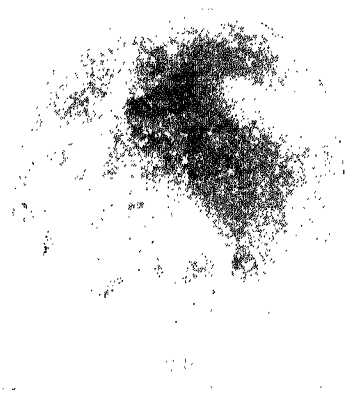
Male crania with trepanned openings.

A. No. 381.

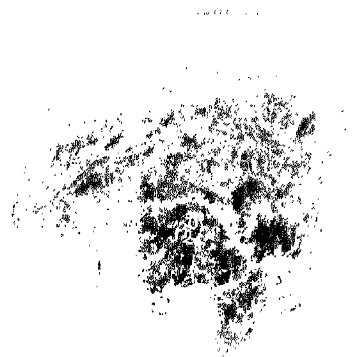
B. No. 673.

Marked artificial deformation of a male (A) and a female (B) cranium.

Risdon: *Skulls from Duweir* (*Lachish*)



A. No. 378.

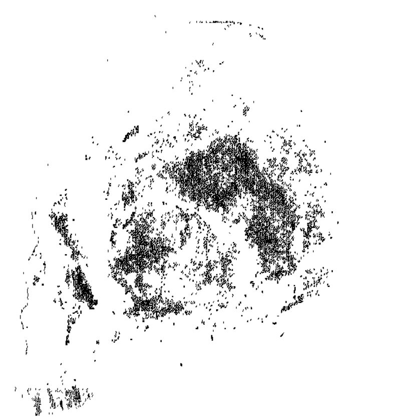


B. No. 375.

C. No. 377.

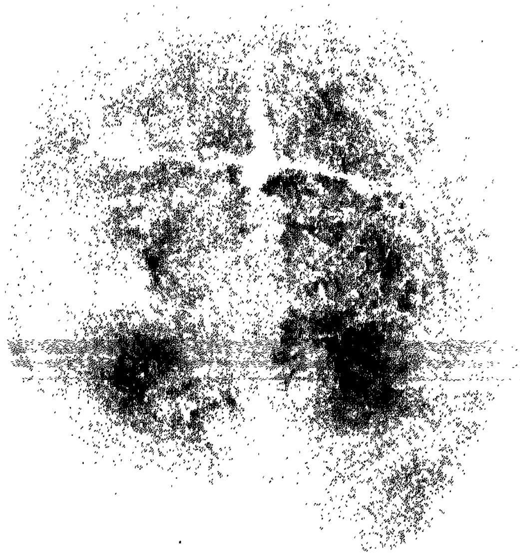
D. No. 376.

E. No. 379.



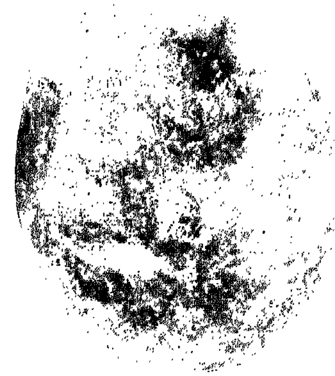
F. No. 380.

ive male crania (A–E) believed to be artificially deformed to a slight extent and a male cranium (F) probably deformed owing to premature closing of the coronal suture.



A. Coronal suture obliterated (No. 380, male).

B. Wound on left of frontal bone (No. 454, female).

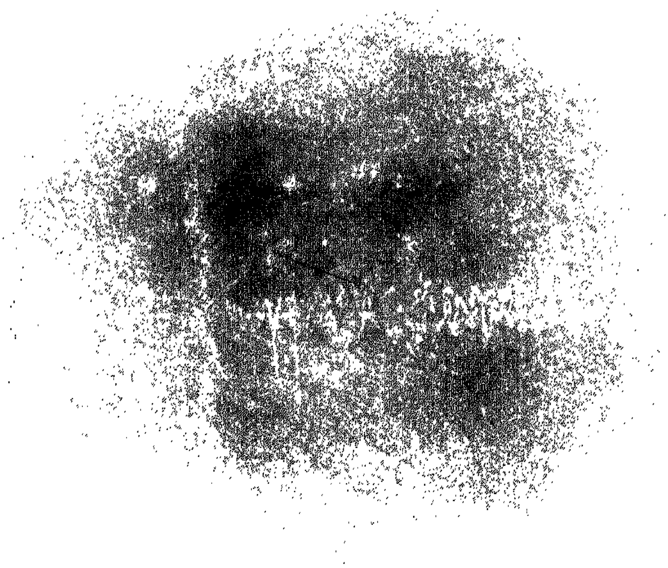


C. Wound on left parietal bone and sagittal suture obliterated (No. 670, female).

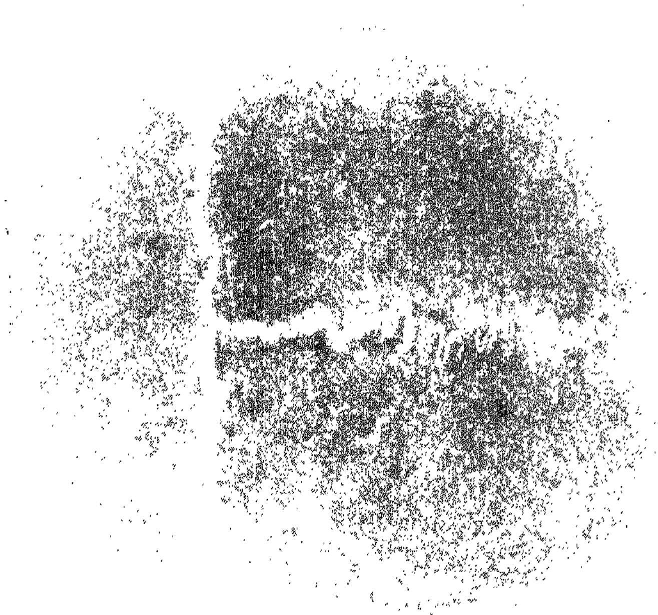
Anomalous crania.

A typical female (No. 388, above) and a typical male cranium (No. 28) of the Lachish series.

A typical female (left) and a typical male cranium of the Lachish series.

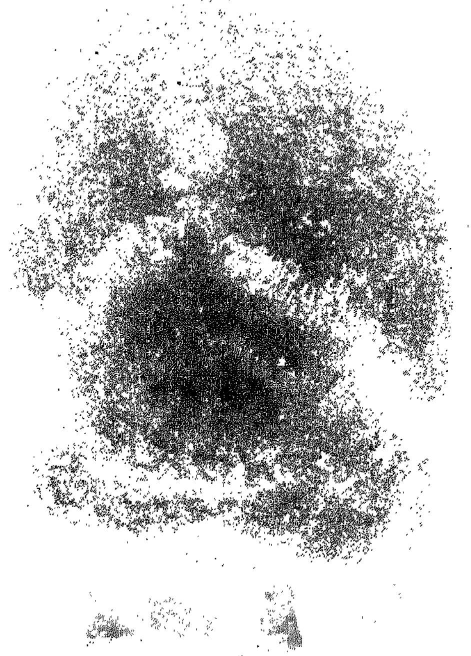


A. No. 388, female.



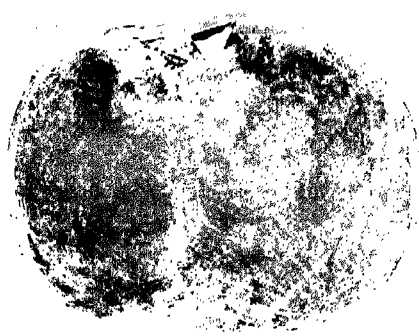
B. No. 28, male.

A typical female and a typical male cranium of the Lachish series.

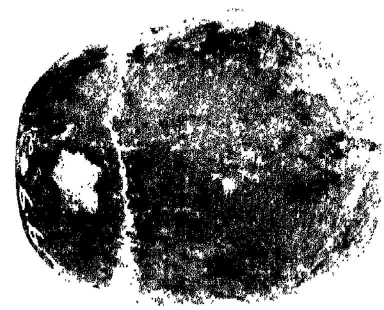


A typical female (No. 388, above) and a typical male cranium (No. 28) of the Lachish series.

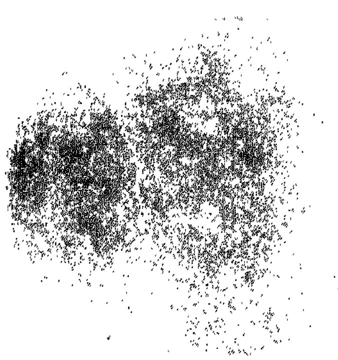
A. No. 672, female.



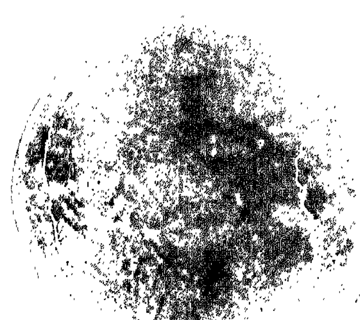
B. No. 364, male.



C. No. 667, female.

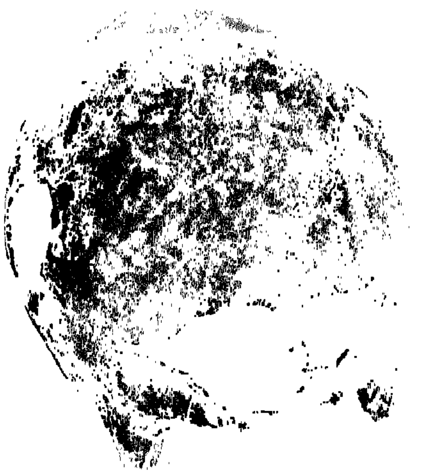


D. No. 366, male.



E. No. 577, female.

Crania showing premature obliteration of the sagittal suture.

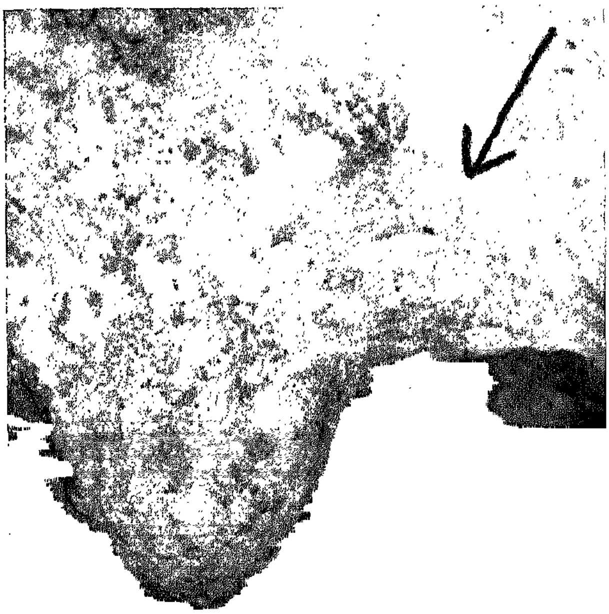


A. No. 299, large wormian bone in right side of coronal suture.

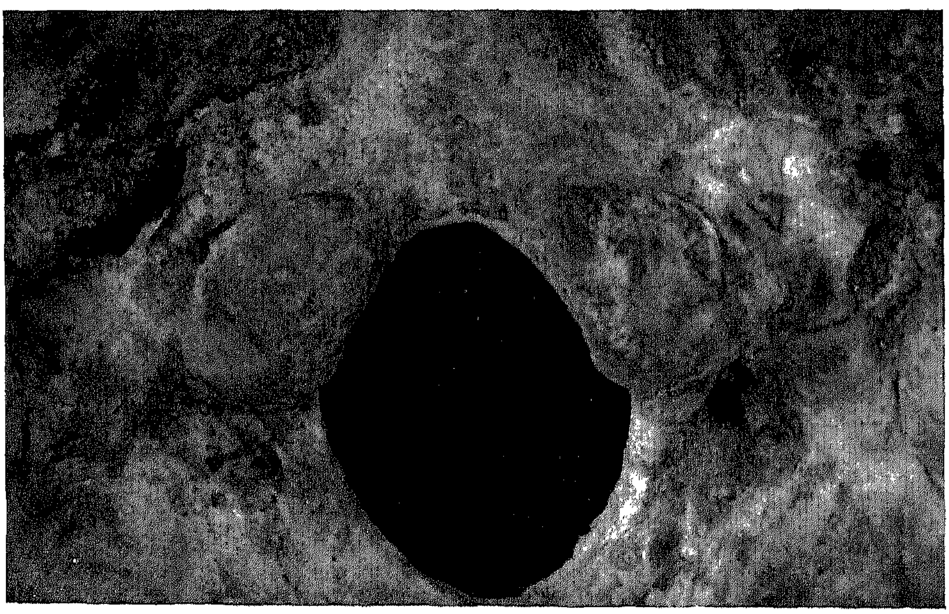
B. No. 60, two symmetrical interparietal bones of unusual form.

C. No. 330, temporal bone largely fused to parietal.

Male crania with sutural anomalies.

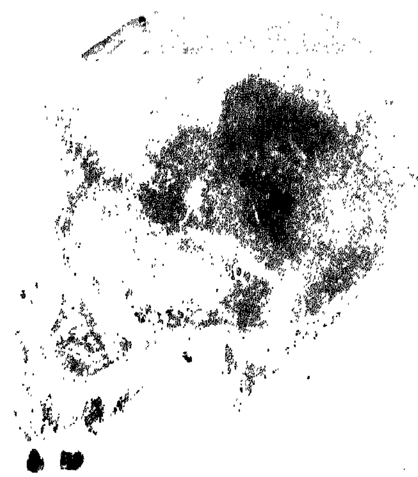


A. No. 324, complete absence of right auricular passage: twice natural size.



B. No. 301, complete absence of left jugular foramen, 1·3 natural size.

Anomalous regions of two male crania.

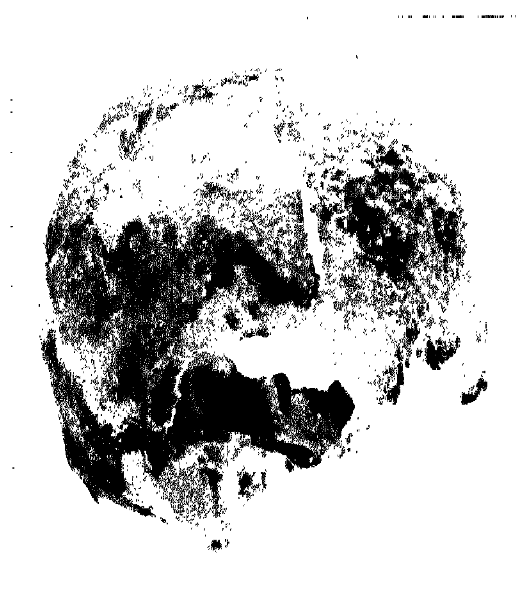


A

B

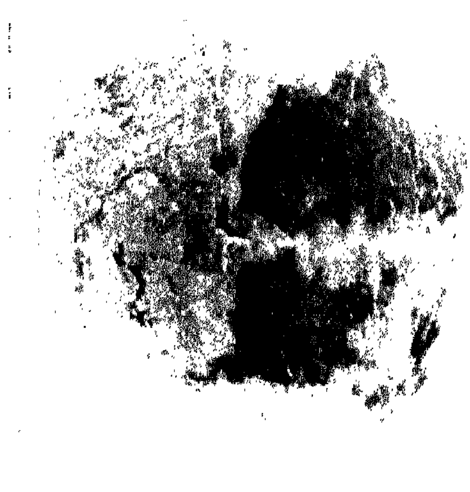
C

A male cranium (No. 382) of unusual form, possibly affected by hydrocephaly.



A

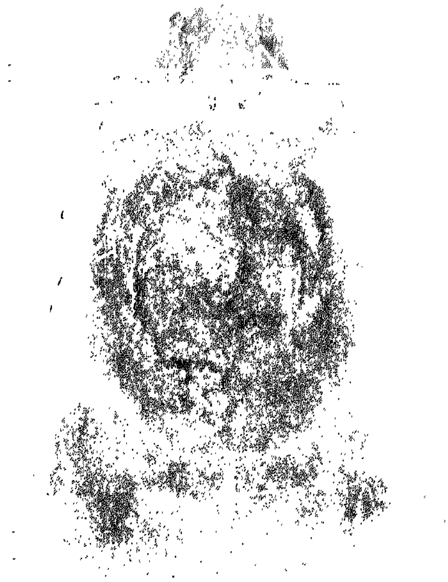
B



C

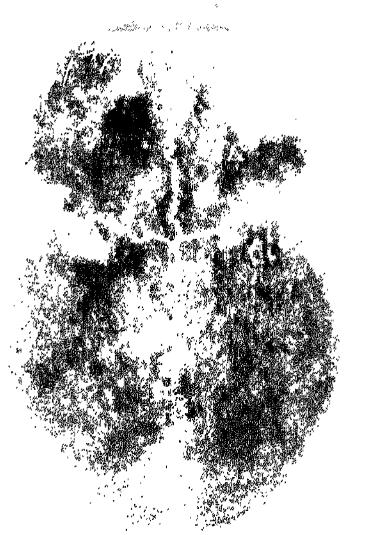
A female cranium (No. 662) with a large diseased area on the frontal bone.

See p. 115 of text.



A. No. 5, male: wound on frontal bone.

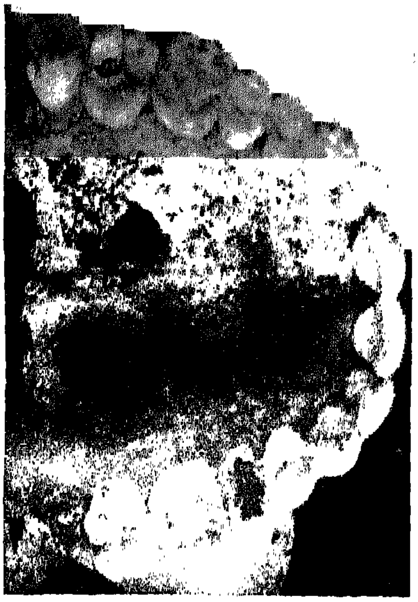
B. No. 1, male: diseased area on right parietal.



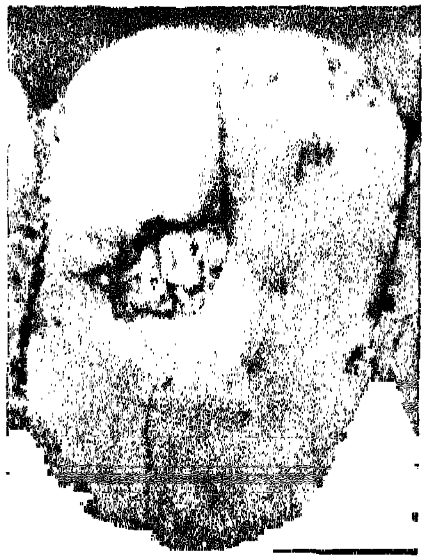
C. No. 419, female: wound on right frontal.

D. No. 513, female: depression in left parietal.

Exceptional crania. See p. 115 of text.



A. The palate: 1·2 natural size.



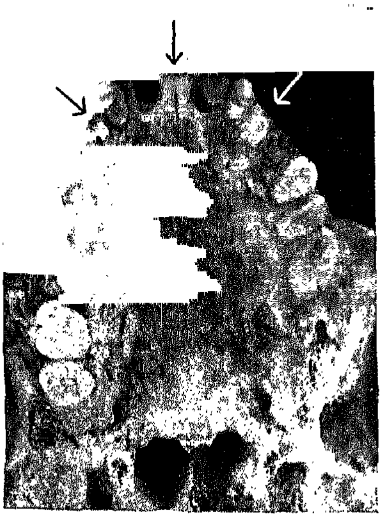
B. The second molar with filling *in situ*: 6·0 natural size.

C. A skiagram showing the filling: 2·7 natural size.

D. Showing the second molar after removal of the filling: 2·7 natural size.

A tooth (upper right second molar) with a filling presumed to be adventitious: No. 518, female.

A. Denticles outside the lower dental arch.



B. Diastemae between central incisors and between lateral incisors and canines.

C. Denticles outside the right side of the lower dental arch.

The anomalous jaws of a female skull (No. 437).

A. Sockets for three incisors only: No. 415, female.

B. Diastemae between canines and premolars and third molars absent: No. 132, male.

C. Supernumerary tooth behind right central incisor: No. 705, juvenile.

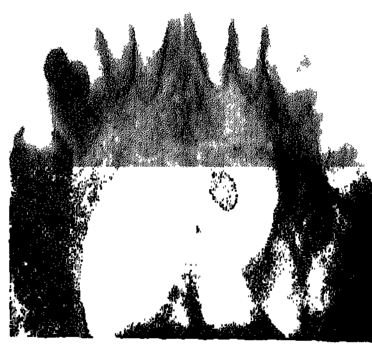
D. Grossly deflected left canine: No. 401, female.

Palates with anomalous dentitions.

A. Incomplete eruption of second and third right molars: No. 506, female.

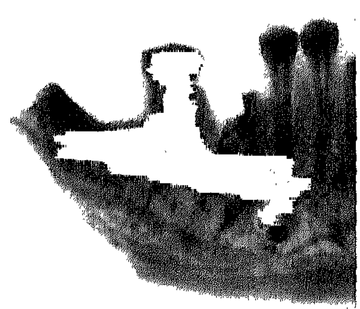
B. Skiagram of the same specimen (A) showing that the roots of the partially erupted teeth were formed.

C. Denticle in palate: No. 383, female.



D. Skiagram of the same specimen (C).

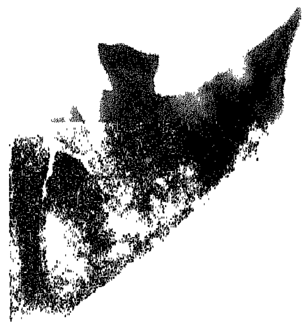
E. Third molar impacted: No. 496, female.



F. Skiagram of the right side of a mandible showing two denticles in the corpus: No. 1056, female.

Jaws with anomalous dentitions.

A. Skiagram of right side of mandible showing the canine and first premolar unerupted: No. [illegible], immature.



B. Skiagram of the left side of the same mandible (A) showing the same teeth on this side unerupted.

C. Third molar in abnormal position and other dental anomalies: No. 72, male. See description on p. 110.

D. Occlusal view of the same jaw (C). The broken roots of the second left molar can be seen.

E. A cyst in the right molar region: No. 469, female.

F. A cyst in the anterior part of the palate: No. 407, female.

Jaws with anomalous dentitions or cysts.

ESTIMATING BACTERIAL POPULATIONS BY THE DILUTION METHOD

By ROBERT DEAN GORDON
Scripps Institution of Oceanography, University of California, La Jolla, California

CONTENTS

	PAGE
Foreword by C. E. ZoBell	167
I. Introduction. Nature and critique of the problem	168
II. Methods of estimating populations from results of successive-dilution technique	170
III. The estimation of log ρ	172
IV. The standard deviation of log ρ	174
V. Transformations. Methods of computing the formulae derived in (III) and (IV)	174
VI. Remarks	178

Foreword by C. E. ZoBell

Several different methods have been described for estimating the densities of bacterial populations in solutions. The most popular of these are plating procedures, the minimum (successive) dilution method, and various direct microscopic counting methods. The accuracy of the two former methods is predicated upon the ability of the bacteria to multiply in nutrient media. Since no one medium under any one set of conditions can provide for the multiplication of bacteria with highly diverse nutritional and environmental requirements, it is not surprising that direct microscopic counts on materials containing a heterogeneous bacterial flora are usually appreciably higher than plate counts or dilution method counts. However, direct microscopic counts fail to differentiate between dead and living bacteria and are beset with almost insurmountable technical difficulties due to the minuteness of the bacteria. Therefore the choice of an enumeration procedure usually rests between plating procedures and the dilution method.

Briefly, and in its simplest form, the dilution method consists of the following procedure: A sample (e.g. 1 c.c.) of the solution under investigation is taken and inoculated into sterile nutrient medium in a test tube. Part of the original solution is then diluted in a certain ratio (usually 10-fold) and the same size sample of this diluted solution is inoculated into nutrient medium in a second test tube. This process is repeated as many times as seems necessary; so that in the end the experimenter will have prepared a series of inocula representing successive 10-fold (or some other ratio) dilutions of the original solution. The highest dilution

in this series which shows growth is then taken to indicate the magnitude of the bacterial density in the original solution.

Such an experiment, however, obviously tells nothing but an integral power of 10 which more or less accurately determines what physicists call the "order of magnitude" of the density. To obtain greater accuracy, investigators (cf. Halvorson & Ziegler, 1933) have suggested that two to ten or more tubes of nutrient medium be inoculated with each dilution of the material. Upon the basis of results obtained in such multiple tube experiments, several methods of estimating the bacterial densities which gave rise to these results have been suggested, as described by Mr Gordon.

In comparing the minimum dilution method with plating procedures for the enumeration of marine bacteria, in this laboratory, the dilution method probability tables of Halvorson & Ziegler (1933) were used. These tables give estimates of bacterial densities based on the numbers of "positives" (i.e. showing growth) observed in ten inocula of each of three successive 10-fold dilutions. More dilutions than this (e.g. five or six) were of course in most cases prepared, and the estimates were based on the most critical set of three of these dilutions.

The results revealed that the dilution method estimates averaged about 20 % higher than plate counts on the same material, although on some individual samples they were actually lower. Duplicate determinations by the two methods on the same samples of material showed that the degree of reproducibility of the plate counts was much higher than that for the dilution method counts. In view of the fact that the dilution method has applications where the plate count cannot be used, further efforts were made to increase its accuracy.

This can be accomplished in either of two ways; namely, by inoculating more tubes with each dilution or by using more dilutions. The latter alternative was tried in which, instead of diluting each time by 10-fold, the dilutions were made $\sqrt{10}$-fold. Under these conditions it seemed that more reliable and more reproducible counts were obtained by the dilution method even when calculated by a crude arithmetical method.

It was at this stage in the experimental work when Mr Gordon was consulted to aid with the calculations. After examining the methods by which Halvorson & Ziegler obtained their results, he expressed the opinion that these methods appeared questionable. He has developed the procedures presented in the following paper for determining the geometric mean estimates, as he has described. It is to be hoped that his method will yield more consistent and more reproducible results. A brief résumé of the results has already been published (Gordon, 1938).

I. Introduction

The underlying assumption by which estimates are made of a bacterial population density, from the results of the successive-dilution technique, is that the individual bacteria are distributed in the space occupied by the fluid medium

containing them, in the same way that we should conceive molecules to be distributed in a solution, if each molecule were unaffected by the presence of any other molecule. (This is not, for instance, true of molecules possessing magnetic or electrical polarity, or of ions.) In other words, the individual bacteria are assumed to be *randomly* distributed in their medium; so that if, for example, a sample of 1 c.c. is taken from 100 c.c. of solution, each bacterium in the whole solution enjoys an *independent* chance of 1 out of 100 (probability $= 0{\cdot}01$) of being caught up in this sample.

Now, in so far as the bacteria exist as individuals they undoubtedly do not satisfy such an assumption "to the letter", but probably exert a certain mutual uniformizing influence on one another. That is, returning to the example just cited, if there are 10,000 bacteria in the 100 c.c. of solution, the probability is likely to be in reality 0, and not $0{\cdot}01^{10,000}$, that all the bacteria should be caught up in the 1 c.c. sample. This, however, simply implies that in a series of such 1 c.c. samples the numbers of bacteria caught up in them are somewhat more closely clustered about the true average density, than we should compute them to be on the basis of the above assumption of complete randomness.

In this sense our assumption of complete randomness amounts simply to "considering a least favourable case" if we are computing the probable "spreads" (standard deviations) of actual counts. On the other hand, the assumption causes us to over-estimate somewhat the probabilities of obtaining no bacteria at all in our 1 c.c. samples, and to underestimate the probabilities of obtaining "one or more" viable bacteria in the samples. It is these probabilities that enter into the estimate of population density by the dilution method. No possibility appears for rationally correcting for such discrepancies; fortunately, however, the two errors committed will to some extent cancel each other in the formulae which follow. But there is suggested in these considerations the possibility of developing rational means of measuring the uniformizing influences associated with various specific cultures in nutrient solution. The values obtained might be correlated with population densities and various specific properties, and might yield very interesting interpretations.

Another, perhaps more serious, objection to the assumption of "randomness", is the known fact that most species of bacteria show tendencies to gather in multicellular groups within the solution, as well as to congregate on glass walls and thus to go out of solution. These processes must affect all methods of counting equally, however, so far as the samples are taken in the same way; hence they should not interfere with the comparability of counts made by different methods.

It might be mentioned here, though, that in comparing plate counts with dilution estimates, account should be taken of the fact that the dilution estimates include obligative anaerobes, while plate counts do not.

II. Methods of estimating populations from results of successive-dilution technique

Assuming random distribution of individuals in a large volume of solution, as described above, let ρ represent the total number of individuals in the whole solution, divided by the total volume occupied. That is, ρ represents the mean density of population in the solution, expressed in a number of individuals per unit volume. Then the probability that in a sample of one unit volume (e.g. 1 c.c.) of solution there will be x individuals (where x is a positive integer) is given by

$$P(x) = \frac{\rho^x}{x!}e^{-\rho}, \tag{1}$$

which is the Poisson distribution.

The probability that there are 0 bacteria is accordingly $P(0) = e^{-\rho}$; and the probability of obtaining one or more bacteria in the sample (i.e. any number except 0) is $1-P(0) = 1-e^{-\rho}$, since the distribution (1) is normalized. Hence if we take ten test-tubes containing sterile nutrient solution, and inoculate each with 1 unit volume of our solution, then the probability that n (≤ 10) of these tubes will have received some number of bacteria, and $(10-n)$ will have received none, is

$$G_\rho(n) = \frac{10!}{n!(10-n)!}(1-e^{-\rho})^n(e^{-\rho})^{10-n}, \tag{2}$$

by well-known rules for computing probabilities. If we understand ρ to represent mean density of *viable* bacteria, then $G_\rho(n)$ in (2) represents likewise the probability that n of the ten test-tubes so inoculated will show growth, and simultaneously the other $(10-n)$ of the tubes will fail to show growth.

Let us now inoculate three rows of ten tubes each, the first row with samples of the original solution, the second row with samples of a 10-fold dilution of the original solution, and the third row with samples of a 100-fold dilution of the original solution. Let ρ represent the mean density of individuals in the middle (10-fold) dilution, and denote by n_{10}, n_1 and $n_{0\cdot1}$ the numbers of tubes showing growth in the three respective rows. Then the probability of observing a result represented by the triplicate $(n_{10}, n_1, n_{0\cdot1}) = n$ is

$$G_\rho^{(3)}(n_{10}, n_1, n_{0\cdot1}) = G_\rho^{(3)}(n) = G'_{10\rho}(n_{10})\,G_\rho(n_1)\,G_{0\cdot1\rho}(n_{0\cdot1})$$
$$= \frac{(10!)^3(e^{-10\rho})^{10-n_{10}}(e^{-\rho})^{10-n_1}(e^{-0\cdot1\rho})^{10-n_{0\cdot1}}(1-e^{-10\rho})^{n_{10}}(1-e^{-\rho})^{n_1}(1-e^{-0\cdot1\rho})^{n_{0\cdot1}}}{n_{10}!\,n_1!\,n_{0\cdot1}!(10-n_{1\cdot0})!\,(10-n_1)!\,(10-n_{0\cdot1})!}. \tag{3}$$

It is upon these three equations (1), (2), and (3) that several methods of estimating population densities from the results $(n_{10}, n_1, n_{0\cdot1}) = n$ of successive dilution experiments are based. (These equations of course refer to successive 10-fold dilutions, and ten tests to each dilution. Analogous equations are easily formed corresponding to other dilution stages and numbers of tests.)

Wells & Wells (1921) have published a method of estimating population density. Their method, applied to the present case, amounts simply to computing three values of ρ, one to fit each of the equations

$$e^{-10\rho} = \frac{10-n_{10}}{10}; \quad e^{-\rho} = \frac{10-n_1}{10}; \quad e^{-0\cdot 1\rho} = \frac{10-n_{0\cdot 1}}{10}, \tag{4}$$

where $(n_{10}, n_1, n_{0\cdot 1})$ are the numbers of "positives" in the three series of tests. The geometric mean $\sqrt[3]{(\rho_{10}\rho_1\rho_{0\cdot 1})}$ of these three values is then accepted as an estimate of the true value of ρ.

On the surface of it, such a way of arriving at an estimate appears rather naive. Such direct computations are usually unsafe when the data are subject to large uncertainties. Apparently the only reason these authors used a geometric mean of the three values instead of some other mean is that from the equation

$$e^{-r\rho} = \frac{10-n_r}{10} \tag{5}$$

we obtain
$$\log\rho = \log\left(\log\frac{10}{10-n_r}\right) - \log r, \tag{6}$$

which is linear in $\log\rho$, $\log r$, and $\log\left(\log\dfrac{10}{10-n_r}\right)$, so that the arithmetic means of these terms are related in the same way as the terms themselves. (Note that $\frac{1}{3}[\log\rho_{10} + \log\rho_1 + \log\rho_{0\cdot 1}] = \log(\rho_{10}\rho_1\rho_{0\cdot 1})^{\frac{1}{3}}$.)

However, bacteria counts are usually expressed "to a certain number (e.g. 3) of significant figures"; that is, interest is centred on the *proportionate* errors, *not* on absolute errors. The difference between no bacteria and one is more important than that between 1000 and 1001. This consideration indicates that the expectation of $\log\rho$, given the data, is more like what we want than, say, the expectation of ρ. But it does not follow that the mean of three separate estimates of $\log\rho$ is the best estimate of $\log\rho$, because the estimates are not equally accurate; they should be weighted by the inverse square of the standard deviation of $\log\rho$ corresponding to an observation n_r, which would have to be computed from an "inverse" distribution derived from (5) by use of Bayes's formula.

Halvorson & Ziegler (1933) published a mimeographed tract in which they presented tables giving the modes of $G_\rho^{(3)}(n)$ (see equation (3)) together with the corresponding maximum values of $G_\rho^{(3)}(n)$, corresponding to various combinations $(n_{10}, n_1, n_{0\cdot 1}) = n$ (that is, considering $G_\rho^{(3)}(n)$ as a function of ρ). Their notation is not the same as that here used; it is as follows:

$$\bar{p} = X = \rho \text{ corresponding to maximum of } G_\rho^{(3)}(n),$$

$$G_{\bar{p}}^{(3)}(n) = G_X^{(3)}(n) = P,$$

$$n_{10} = p_1; \quad n_1 = p_2; \quad n_{0\cdot 1} = p_3.$$

This estimate $\bar{p}$ of ρ is obviously that obtained by the "method of maximum likelihood", which is associated with the name of R. A. Fisher. The relation of

this method to the principle of inverse probability has been discussed by Jeffreys (1938*a*, *b*). It is equivalent to taking a uniform prior probability for ρ, and then adopting the mode of the posterior probability as the estimate.

In the present instance the significance of a modal estimate of ρ is reduced by the fact that the *a posteriori* distributions of ρ (computed in particular from uniform *a priori* probability) vary widely in skewness with different values of the n_r's. This is indicated by the comparison of three computed geometric means with corresponding modal estimates in the last section of this paper. The point is verified by a series of laboratory comparisons of modal (Halvorson & Ziegler) estimates with direct plate counts on the same materials, reported elsewhere (Gordon & ZoBell, 1939). Thus the relationship of the mode to the distribution of which it is a representative is a varying quality, and the mode is deprived of any immediate significance.

It has already been mentioned that the particular value of ρ that is best adapted to the purposes of bacteriology corresponds to the expectation of the logarithm, on the ground that we are more interested in the ratio of ρ in two samples than in the absolute difference. Further, since ρ can have values from 0 to ∞, but $\log \rho$ from $-\infty$ to $+\infty$, the probability of $\log \rho$ has an opportunity of being nearly normally distributed that is denied to ρ; similarly for correlation $\tanh^{-1}\rho$, and for the ratio of two standard deviations $\log(s_1/s_2)$, have nearly normal probability distributions, and have been extensively used for that reason on Fisher's recommendation. If we can find the expectation of $\log \rho$ and the second moment of its probability distribution, a normal curve with corresponding mean and standard deviation should give a good representation of the law as a whole.

III. The estimation of log ρ

Using the Bayes-Laplace formula with uniform prior probability for ρ, we obtain from (3) for the posterior probability distribution of ρ,

$$P(\rho) = \frac{G_\rho^{(3)}(n)}{\int_0^\infty G_\rho^{(3)}(n)\,d\rho}. \tag{7}$$

The expectation of $\log \rho$ is

$$\log \bar{\rho} = \int_0^\infty \log \rho \; P(\rho)\,d\rho,$$

which, cancelling factors in (3) and (7), equals

$$\log \bar{\rho} = \frac{\int_0^\infty \log \rho\, e^{-(111-10n_{10}-n_1-0\cdot 1n_{0\cdot 1})\rho}\,(1-e^{-10\rho})^{n_{10}}\,(1-e^{-\rho})^{n_1}\,(1-e^{-0\cdot 1\rho})^{n_{0\cdot 1}}\,d\rho}{\int_0^\infty e^{-(111-10n_{10}-n_1-0\cdot 1n_{0\cdot 1})\rho}\,(1-e^{-10\rho})^{n_{10}}\,(1-e^{-\rho})^{n_1}\,(1-e^{-0\cdot 1\rho})^{n_{0\cdot 1}}\,d\rho}, \tag{8}$$

The integrands can obviously be expressed in the form

$$[\log \rho \text{ or } 1](1-E^{100})^{n_{10}}(1-E^{10})^{n_1}(1-E)^{n_{0\cdot1}}e^{-\frac{1}{10}x_0\rho},$$

where E is the Boolean operator $=(1+\Delta)$, Δ meaning difference with regard to

$$x_0 = 1110-100n_{10}-10n_1-n_{0\cdot1}.$$

Since E is commutative with the sign of integration, we may therefore write

$$\log \bar{\rho} = \frac{(1-E^{100})^{n_{10}}(1-E^{10})^{n_1}(1-E)^{n_{0\cdot1}}\int_0^\infty \log \rho \, e^{-\frac{1}{10}x_0\rho}\,d\rho}{(1-E^{100})^{n_{10}}(1-E^{10})^{n_1}(1-E)^{n_{0\cdot1}}\int_0^\infty e^{-\frac{1}{10}x_0\rho}\,d\rho}. \tag{9}$$

That this is so may be seen by expanding the operator

$$H = (1-E^{100})^{n_{10}}(1-E^{10})^{n_1}(1-E)^{n_{0\cdot1}} \tag{10}$$

as a polynomial in E, then operating term by term on the integrands. The result will be the same as the result of expanding the integrands of (8) in powers of e.

The integrals in (9) may be determined by means of the following identities:

$$\left.\begin{aligned}\int_0^\infty e^{-x\rho}\rho^u\,d\rho &= \frac{\Gamma(u+1)}{x^{u+1}},\\ \int_0^\infty \log\rho\, e^{-x\rho}\rho^u\,d\rho &= \frac{d}{du}\left(\frac{\Gamma(u+1)}{x^{u+1}}\right)\\ &= \frac{\Gamma(u+1)}{x^{u+1}}(F(u+1)-\log x),\end{aligned}\right\} \tag{11}$$

where $F(u+1)$ is the "digamma function" defined by $F(u+1) = d\log\Gamma(u+1)/du$. (see *British Association Tables*, vol. 1 (1931) with a different notation).

Place $u=0$ in these identities; there results

$$\left.\begin{aligned}\int_0^\infty e^{-\frac{1}{10}x_0\rho}\,d\rho &= \frac{10}{x_0},\\ \int_0^\infty \log\rho\, e^{-\frac{1}{10}x_0\rho}\,d\rho &= \frac{10}{x_0}\left(F(1)-\log\frac{x_0}{10}\right),\end{aligned}\right\} \tag{12}$$

where $$F(1) = -0\cdot57722.$$

In this way we obtain simply from (9), for $\log\bar{\rho}$

$$\log\bar{\rho} = F(1)+\log 10 - \frac{H\left\{\dfrac{\log x_0}{x_0}\right\}}{H\left\{\dfrac{1}{x_0}\right\}},\ldots \tag{13}$$

where H is the operator defined in (10).

IV. The standard deviation of $\log p$

The standard deviation of $\log p$ is the square root of the second moment of $\log p$ about its mean $\log \bar{p}$. Using the operator-notation of the previous section, with standard formulae of statistics, we obtain

$$\sigma^2_{\log p} = \frac{H\left(\int_0^\infty (\log p)^2 e^{-\frac{1}{10}x_0 p}\, dp\right)}{H\left(\int_0^\infty e^{-\frac{1}{10}x_0 p}\, dp\right)} - (\log \bar{p})^2. \tag{14}$$

Now, as before,

$$\begin{aligned}\int_0^\infty (\log p)^2 e^{-\frac{1}{10}x_0 p}\, dp &= \left[\frac{d^2}{du^2}\int_0^\infty p^u e^{-\frac{1}{10}x_0 p}\, dp\right]_{u=0} \\ &= \frac{10}{x_0}\left[\{F(1)\}^2 + F'(1) + \{\log 10\}^2 + 2F(1)\log 10\right. \\ &\qquad \left. -2\{\log 10 + F(1)\}\log x_0 + \{\log x_0\}^2\right],\end{aligned} \tag{15}$$

whence we easily find

$$\begin{aligned}\sigma^2_{\log p} = &-(\log \bar{p})^2 + [F(1) + \log 10]^2 + F'(1) \\ &- 2[F(1) + \log 10]\frac{H\left\{\frac{\log x_0}{x_0}\right\}}{H\left\{\frac{1}{x_0}\right\}} \\ &+ \frac{H\left\{\frac{(\log x_0)^2}{x_0}\right\}}{H\left\{\frac{1}{x_0}\right\}},\end{aligned} \tag{16}$$

where $\quad F(1) = -0{\cdot}577216, \quad F'(1) = 1{\cdot}644934.$

V. Transformations

In the fractional terms in (13) and (16), the numerators and denominators each represent finite alternating series; but these series are useless for computing purposes, because the terms are so nearly equal that each would have to be expressed with at least twenty significant figures to yield sufficiently accurate results. Hence we carry out the following transformations:

Referring to (10) we take out $(1-E)^{\Sigma n} = (-\Delta)^{\Sigma n}$ as a factor from the operator H, and obtain

$$H = H_0(-\Delta)^N, \tag{17}$$

where $N = \Sigma n = n_{10} + n_1 + n_{0\cdot 1}$, and

$$H_0 = (1 + E + \ldots + E^{99})^{n_{10}}(1 + E + \ldots + E^9)^{n_1}. \tag{18}$$

If this last expression is expanded, evidently all terms must turn out to have positive coefficients; so that if the functions

$$\Delta^N\left(\frac{1}{x}\right), \quad \Delta^N\left(\frac{\log x}{x}\right), \quad \Delta^N\left(\frac{(\log x)^2}{x}\right),$$

can be accurately determined, the required end result is simply a weighted sum of terms of like sign, and there is no cancellation.

In the first place it is very simple to obtain

$$(-\Delta)^N\left(\frac{1}{x}\right) = \frac{N!(x-1)!}{(x+N)!} \tag{19}$$

by mathematical induction.

To obtain $(-\Delta)^N(\log x/x)$ it suffices to expand $\log(x+v)/(x+v)$ as a power series in v and operate with $(-\Delta)^N$ termwise on successive powers of $v=0$, using the formula

$$\Delta^N 0^p = \frac{p!}{(p-N)!}B_{p-N}^{(-N)}, \tag{20}$$

where $B_{p-N}^{(-N)}$ are the "generalized Bernoulli numbers" of order $(-N)$ and degree $(p-N)$ (cf. L. M. Milne-Thomson, 1933, p. 134). By this means we finally obtain

$$(-1)^N\Delta^N\frac{\log x}{x} = \frac{N!(x-1)!}{(x+N)!}\log x + \sum_{p=N}^{\infty}\frac{(-1)^{p-N+1}}{x^{p+1}}\frac{p!}{(p-N)!}\left(1+\frac{1}{2}+\ldots+\frac{1}{p}\right)B_{p-N}^{(-N)}. \tag{21}$$

By similarly treating $\log(x+N-v)/(x+N-v)$ it is easy to obtain also

$$(-1)^N\Delta^N\frac{\log x}{x} = \frac{N!(x-1)!}{(x+N)!}\log(x+N) - \sum_{p=N}^{\infty}\frac{1}{(x+N)^{p+1}}\frac{p!}{(p-N)!}\left(1+\frac{1}{2}+\ldots+\frac{1}{p}\right)B_{p-N}^{(-N)}. \tag{21a}$$

These formulae (21) and (21 a) can evidently both be used simultaneously to obtain upper and lower bounds to the required result. Both are convergent if $x>N$.

The sums
$$1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{p}$$

in (21) and (21 a) can be computed directly, or by means of various short formulae of which the following may be suggested (Boole, 1872, p. 92):

$$1+\frac{1}{2}+\ldots+\frac{1}{p} = 0{\cdot}577216+\log p+\frac{1}{2p}-\frac{1}{12p^2}+\frac{1}{120p^4}\cdots \tag{22}$$

The Bernoulli numbers are easily computed by means of the reversion formula (Milne-Thomson, 1933, p. 129),

$$B_{\upsilon}^{(-N)} = \frac{\upsilon B_{\upsilon-1}^{(-N)} + B_{\upsilon}^{(-N+1)}}{1+\dfrac{\upsilon}{N}}, \tag{23}$$

together with the values $B_{\upsilon}^{(0)} = 0$ if $\upsilon \neq 0$; $B_0^{(-N)} = 1$. Also polynomial formulae are given in Davis's tables. We have computed them as far as $N = 27$, $\upsilon = 7$, of which a tabulation is presented at the end of this paper.

To deal with the function $\Delta^N \left\{ \dfrac{(\log x)^2}{x} \right\}$,

we make use of Boole's operator identity

$$\Delta_x^N = (e^{d/dx} - 1)^N. \tag{24}$$

If this is expanded in powers of d/dx and made to operate on $(\log x)^2/x$, the result is a series in the derivatives of the latter function. These derivatives are found to have the form

$$\frac{d^p}{dx^p}\frac{(\log x)^2}{x} = \frac{(-1)^p p!(\log x)^2}{x^{p+1}} + \frac{p}{x^{p+1}}\{2\log x\, B_{p-1}^{(p+1)} + (p-1)\, B_{p-2}^{(p+1)}\},$$

and the final result is

$$\Delta^N \frac{(\log x)^2}{x} = \sum_{p=N}^{\infty} \frac{p!}{(p-N)!} \frac{B_{p-N}^{(-N)}}{x^{p+1}} \left\{ (-1)^p (\log x)^2 + \frac{2\log x\, B_{p-1}^{(p+1)}}{(p-1)!} + \frac{B_{p-2}^{(p+1)}}{(p-2)!} \right\}. \tag{25}$$

In an analogous manner we also obtain

$$\Delta^N \frac{(\log x)^2}{x} = (-1)^N \sum_{p=N}^{\infty} \frac{p!}{(p-N)!} \frac{B_{p-N}^{(-N)}}{(x+N)^{p+1}} \left\{ [\log(x+N)]^2 + \frac{(-1)^p\, 2\log(x+N)\, B_{p-1}^{(p+1)}}{(p-1)!} + \frac{(-1)^p\, B_{p-2}^{(p+1)}}{(p-2)!} \right\}. \tag{25a}$$

As (21) and (21*a*), so also (25) and (25*a*), can always serve together to give upper and lower bounds to the values sought, so long as $x_0 > N$.

For use in the last equations, the following formulae may be found useful:

$$\frac{B_{p-1}^{(p+1)}}{(p-1)!} = (-1)^{p-1}\left(1+\frac{1}{2}+\frac{1}{3}+\dots+\frac{1}{p}\right), \tag{26}$$

$$\frac{B_{p-2}^{(p+1)}}{(p-2)!} = (-1)^p \sum_{r=1}^{p-1} \frac{1}{r} \sum_{s=1}^{p-r} \frac{(-1)^{s-1}}{s}. \tag{27}$$

(i) *Expansion of the operator H_0*

The operator H_0 (equation (18)) taken as a polynomial in E, has the form of

$$\begin{aligned} \mathfrak{H}_0(z) &= (1+z+\dots+z^{99})^{n_{10}}(1+z+\dots+z^9)^{n_1} \\ &= \frac{(1-z^{100})^{n_{10}}(1-z^{10})^{n_1}}{(1-z)^{n_{10}+n_1}} \\ &= P_0 + P_1 z + P_2 z^2 + \dots + P_{\upsilon} z^{\upsilon} + \dots + P_{99n_{10}+9n_1} z^{99n_{10}+9n_1}, \end{aligned}$$

where $$P_v = P_{99n_{10}+9n_1-v} = \sum_{r=0}^{\cdot 01v} \sum_{s=0}^{\cdot 1v-10r} (-1)^{r+s} C_r^{n_{10}} C_s^{n_1} C_{v-100r-10s}^{v-100r-10s+n_{10}+n_1-1}, \tag{28}$$

with $$C_a^b = \frac{b!}{a!(b-a)!}.$$

Thus, with these values of the P_v's, we have

$$H_0 = P_0 + P_1 E + P_2 E^2 + \ldots + P_v E^v + \ldots. \tag{29}$$

(ii) *Computation of P_v for large v*

When v is fairly large, that is, considerably different from both 0 and $99n_{10}+9n_1+1$, the formula (28) becomes unwieldy and inaccurate. By means of a Laplace transformation, that is, by making use of the so-called "characteristic function" for P_v (see Uspensky, 1937, chapter XII), we are able to obtain the following Fourier series representing the continuous function which is obtained by plotting the points (v, P_v) and then connecting them with straight lines:

$$P_v = P_{99n_{10}+9n_1-v} = \sum_{\mu=0}^{\infty} A_\mu \cos \frac{2\pi\mu}{K}(v-b), \tag{30}$$

where $$K = 99n_{10}+9n_1+2; \quad b = \tfrac{1}{2}(99n_{10}+9n_1);$$

and $$A_0 = \frac{10^{2n_{10}+n_1}}{99n_{10}+9n_1+2}; \quad A_\mu = \frac{K}{8\pi^2\mu^2} \frac{\left(\sin\frac{100\pi\mu}{K}\right)^{n_{10}} \left(\sin\frac{10\pi\mu}{K}\right)^{n_1}}{\left(\sin\frac{\pi\mu}{K}\right)^{n_{10}+n_1-2}}$$

for $\mu \neq 0$.

If μ/K is an integer, $\neq 0$, then

$$A_\mu = \frac{K}{8\pi^2\mu^2} \frac{0^{n_{10}+n_1}}{0^{n_{10}+n_1-2}} = 0.$$

Otherwise obviously

$$|A_\mu| < \frac{10^{2n_{10}+n_1}K}{8\pi^2\mu^2}\left(\sin\frac{\pi\mu}{K}\right)^2$$

$$< \frac{10^{2n_{10}+n_1}K}{8\pi^2\mu^2}\left(\sin\frac{\pi}{200}\right)^2 \text{ if } |\mu/K - n| < \frac{1}{200}, \; n = \text{integer};$$

$$|A_\mu| < \frac{10^{n_1}K}{8\pi^2\mu^2}\left(\sin\frac{\pi}{200}\right)^{-n_{10}+2} \text{ if } \frac{1}{200} \leqq |\mu/K - n| < \frac{1}{20};$$

$$|A_\mu| < \frac{K}{8\pi^2\mu^2}\left(\sin\frac{\pi}{20}\right)^{-n_{10}-n_1+2}, \text{ if } |\mu/K - n| \geqq \frac{1}{20}.$$

The first limit is the greatest. It follows that

$$\sum_{\mu=p+1}^{\infty} A_\mu \cos\frac{2\pi\mu}{K}(\xi - b) < \sum_{\mu=p+1}^{\infty} |A_\mu| < \frac{0{\cdot}000247 \times 10^{2n_{10}+n_1}K}{8\pi^2} \sum_{\mu=p+1}^{\infty} \frac{1}{\mu^2}$$

$$< \frac{0{\cdot}000247 \times 10^{2n_{10}+n_1}K}{8\pi^2}\int_p^\infty \frac{dx}{x^2} < \frac{0{\cdot}000247 \times 10^{2n_{10}+n_1}K}{8\pi^2 p} \tag{31}$$

which can serve as an error function for (30).

VI. Remarks

Sufficient mathematical tools are presented in equations (19), (21), (21*a*), (25), (25*a*), (28), and (30), for actually computing the geometric mean estimates $\bar{\rho}$ of densities of bacteria from the results of successive dilution experiments, together with the standard deviations of their natural logarithms, from formulae (13) and (16). The standard deviation of the logarithm of course will be a measure of the proportionate accuracy of the estimate, in the same sense that the ordinary standard deviation is a measure of the arithmetical accuracy. (It may be remarked, incidentally, that *all logarithms occurring in the above formulae are understood to be natural logarithms.*)

If we assume in general that the function $P(\rho)$ defined in (7) can be very nearly approximated by a normal distribution function with respect to $\log \rho$—which should be satisfactory for bacteriological purposes—then corresponding to $\sigma_{\log \rho}$ there will be a "probable error" $0{\cdot}675\sigma_{\log \rho} = s$. To this will correspond a "probable error ratio"

$$\epsilon = e^s - 1 = 10^{0\cdot293\sigma} - 1,$$

that is
$$\log_{10}(1+\epsilon) = 0{\cdot}293\sigma_{\log \rho}.$$

The meaning of ϵ is that the probability is approximately $\frac{1}{2}$, that

$$-\frac{\epsilon}{1+\epsilon}\bar{\rho} < \rho - \bar{\rho} < \epsilon\bar{\rho}.$$

To save labour, it remains yet to determine limits to the extents of errors committed by adding only, say, every twenty-fifth term corresponding to the operator

$$H_0 = \sum_{(\nu)} P_\nu E^\nu,$$

in making the computations; or else to determine short formulae for the required sums. For the purposes of rigour, upper bounds will also have to be determined for the Bernoulli numbers used in the formulae.

Mrs Naomi Lancaster has made rough computations of several values of $\bar{\rho}$ as shown below:

Argument			$\bar{\rho}$ computed by us	$\bar{\rho}$ from Halvorson & Ziegler	% deviation from Halvorson & Ziegler
n_{10}	n_1	$n_{0\cdot1}$			
10	7	3	1·43	1·53	−7·0%
8	5	1	0·291	0·267	+9·0%
4	2	1	0·086	0·080	+7·5%

If we were to plot these percentage deviations against n_{10}, we should be led to expect a maximum positive deviation of perhaps 12 % or more in the region of $n_{10} = 6$ or 7.

Dr ZoBell has compared plate counts with corresponding estimates from Halvorson & Ziegler's tables, and their ratios show exactly similar trends to those indicated by the above three comparisons, but relatively even more pronounced. These will probably be reported elsewhere in detail.

In order to make possible the practical use of these results by bacteriologists and others, as well as to test their validity experimentally, it will be necessary to prepare tables of $\bar{p}$ and ϵ as formulated in the preceding pages, corresponding to the tables previously prepared by Halvorson & Ziegler. This will require financial aid from some source.

The writer wishes to express his appreciation to Dr C. E. ZoBell for acquainting him with this very interesting problem, and to Dr George F. McEwen for his encouragement and occasional aid in carrying out the work and making available the necessary literature. Dr Harry Bateman of the Mathematics Department, California Institute of Technology, and Dr Risselman of the University of California at Los Angeles, were also helpful in bringing the analysis through one or two difficult points.

APPENDIX

Table of Bernoulli numbers $B_v^{(-N)}$

$v=$	1	2	3	4	5	6	7
$N=1$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{7}$	$\frac{1}{8}$
2	1	$\frac{7}{6}$	$\frac{3}{2}$	$\frac{31}{15}$	3	$\frac{127}{28}$	$\frac{85}{12}$
3	$\frac{3}{2}$	$\frac{5}{2}$	$\frac{9}{2}$	$\frac{43}{5}$	$\frac{69}{4}$	$\frac{3025}{84}$	$\frac{311}{4}$
4	2	$\frac{13}{3}$	10	$\frac{243}{10}$	$\frac{185}{3}$	$\frac{6821}{42}$	$\frac{1325}{3}$
5	$\frac{5}{2}$	$\frac{20}{3}$	$\frac{75}{4}$	$\frac{331}{6}$	$\frac{675}{4}$	$\frac{11215}{21}$	$\frac{5225}{3}$
6	3	$\frac{19}{2}$	$\frac{63}{2}$	$\frac{1087}{10}$	$\frac{777}{2}$	$\frac{30083}{21}$	5432
7	$\frac{7}{2}$	$\frac{77}{6}$	49	$\frac{1939}{10}$	$\frac{4753}{6}$	$\frac{9992}{3}$	$\frac{43120}{3}$
8	4	$\frac{50}{3}$	72	$\frac{4819}{15}$	1476	$\frac{146240}{21}$	33664
9	$\frac{9}{2}$	21	$\frac{405}{4}$	$\frac{2514}{5}$	2565	$\frac{93886}{7}$	$\frac{573975}{8}$
10	5	$\frac{155}{6}$	$\frac{275}{2}$	752	$\frac{12650}{3}$	$\frac{677465}{28}$	$\frac{567325}{4}$
11	$\frac{11}{2}$	$\frac{187}{6}$	$\frac{363}{2}$	$\frac{16258}{15}$	$\frac{26499}{4}$	$\frac{1158509}{28}$	$\frac{3164029}{12}$

Table of Bernoulli numbers $B_v^{(-N)}$ (cont.)

$v=$	1	2	3	4	5	6	7
$N=12$	6	37	234	$\frac{15149}{10}$	10023	$\frac{2842373}{42}$	465725
13	$\frac{13}{2}$	$\frac{130}{3}$	$\frac{1183}{4}$	$\frac{20631}{10}$	$\frac{176267}{12}$	$\frac{2238730}{21}$	$\frac{9453353}{12}$
14	7	$\frac{301}{6}$	$\frac{735}{2}$	$\frac{82439}{30}$	$\frac{41895}{2}$	$\frac{975623}{6}$	$\frac{7704025}{6}$
15	$\frac{15}{2}$	$\frac{115}{2}$	450	$\frac{7181}{2}$	29175	$\frac{10129615}{42}$	2026550
16	8	$\frac{196}{3}$	544	$\frac{23066}{5}$	$\frac{119408}{3}$	$\frac{7330868}{21}$	$\frac{9329056}{3}$
17	$\frac{17}{2}$	$\frac{221}{3}$	$\frac{2601}{4}$	$\frac{87601}{15}$	$\frac{106641}{2}$	$\frac{10384229}{21}$	$\frac{37236205}{8}$
18	9	$\frac{165}{2}$	$\frac{1539}{2}$	$\frac{36483}{5}$	70281	$\frac{19239635}{28}$	$\frac{27257571}{4}$
19	$\frac{19}{2}$	$\frac{551}{6}$	$\frac{1805}{2}$	$\frac{45049}{5}$	$\frac{1095635}{12}$	$\frac{26279679}{28}$	$\frac{39123375}{4}$
20	10	$\frac{305}{3}$	1050	$\frac{66049}{6}$	117075	$\frac{17672415}{14}$	$\frac{41371225}{3}$
21	$\frac{21}{2}$	112	$\frac{4851}{4}$	$\frac{133217}{10}$	$\frac{7714707}{52}$	$\frac{5022439}{3}$	$\frac{38264149}{2}$
22	11	$\frac{737}{6}$	$\frac{2783}{2}$	$\frac{159819}{10}$	$\frac{1115983}{6}$	$\frac{46037134}{21}$	$\frac{78466685}{3}$
23	$\frac{23}{2}$	$\frac{805}{6}$	1587	$\frac{570653}{30}$	$\frac{461817}{2}$	$\frac{59587135}{21}$	$\frac{105841262}{3}$
24	12	146	1800	$\frac{112379}{5}$	284100	$\frac{76306988}{21}$	47006000
25	$\frac{25}{2}$	$\frac{475}{3}$	$\frac{8125}{4}$	26380	$\frac{1040000}{3}$	$\frac{96763700}{21}$	$\frac{1486135625}{24}$
26	13	$\frac{1027}{6}$	$\frac{4563}{2}$	$\frac{461578}{15}$	419796	$\frac{486388487}{84}$	$\frac{322887513}{4}$
27	$\frac{27}{2}$	$\frac{369}{2}$	$\frac{5103}{2}$	$\frac{178452}{5}$	$\frac{2019087}{4}$	$\frac{202034577}{28}$	$\frac{416849895}{4}$

REFERENCES

Boole, G. (1872). *Finite Differences.* London: Macmillan.

Gordon, R. D. (1938). *Proc. Nat. Acad. Sci., Wash.*, **24**, 212–15.

Gordon, R. D. & ZoBell, C. E. (1939). *Zbl. Bakt.* **2** (in the Press).

Halvorson, H. O. & Ziegler, N. R. (1933). *Quantitative Bacteriology.* Minneapolis: Burgess.

Jeffreys, H. (1938*a*). *Ann. Eugen., Lond.*, **8**, 146–51.

—— (1938*b*). *Proc. Roy. Soc.* A, **167**, 464–83.

Milne-Thomson, L. M. (1933). *Calculus of Finite Differences.* London: Macmillan.

Uspensky, J. V. (1937). *Introduction to Mathematical Probability.* New York: McGraw-Hill.

Wells, P. V. & Wells, W. F. (1921). *J. Wash. Acad. Sci.* **11**, 265–73.

NOTE ON THE INVERSE AND DIRECT METHODS OF ESTIMATION IN R. D. GORDON'S PROBLEM

By E. S. PEARSON

THE practical solution of the problem of estimating the mean density of a bacterial population by the dilution method requires not only the determination of a single-valued estimate, but also some measure of the reliability of this estimate. By the introduction of an ingenious mathematical procedure Mr Gordon has taken, in the preceding paper, a first step in the solution of this problem on lines involving the application of Bayes's theorem.* Thus if ρ is the mean density per unit volume at a given dilution, he obtains an *a posteriori* distribution for ρ and, since it seems likely that the derived distribution of $\log \rho$, rather than that of ρ, will be approximately normal, he shows how the expectation and standard deviation of $\log \rho$ may be calculated. With the help of these a probability statement regarding the unknown ρ of the form given on p. 178 may be made. To put the results into working form extensive computation will, however, be necessary.

Contrasted with this inverse approach is what may be termed the direct solution involving the determination of a fiducial or confidence interval. This solution requires (*a*) the choice of an appropriate sample estimate of ρ, say R, and (*b*) the determination of its sampling distribution, say $p(R \mid \rho)$. Mr Gordon refers to R. A. Fisher's maximum likelihood estimate, say R_L (Fisher, 1922, pp. 363–5), whose calculation in the case where ten tubes are examined at each of three dilution levels, the dilution factor being 10 : 1, has been made easy by the tables of Halvorson & Ziegler (1933*a*), but convinced as he clearly is that the inverse approach is the only legitimate one, he no doubt did not feel it necessary to discuss the possibility of a fiducial solution as alternative to his own. Since, however, a paper by Matuszewski *et al.* (1935) did present a preliminary working solution of this kind for just the same experimental arrangement—I will call it the 10, 3, 10 arrangement—discussed by Halvorson & Ziegler and by Gordon, it may be useful to make some reference to this result, and also to add a few comments on the difference between the two lines of approach.

In their first paper (1933*a*, p. 121) Halvorson & Ziegler gave a table showing the value of R_L corresponding to each (or rather, most) of the possible combinations of n_{10}, n_1, and $n_{0\cdot1}$, the number of tubes (in Gordon's notation) out of the ten tested at each dilution which show growth. In their third paper (1933*c*) they carried out some investigations into the sampling variation of R_L for a fixed ρ.

* It is interesting to note that the pioneer paper on this subject by Greenwood & Yule (1917) also followed the Bayes's theorem approach.

Taking the 10, 3, 10 arrangement and the four cases $\rho = 0{\cdot}15$, $0{\cdot}25$, $0{\cdot}50$ and $1{\cdot}50$, they calculated from the term of an appropriate multinomial expansion the probability of observing different combinations of n_{10}, n_1, and $n_{0\cdot1}$; hence, using their tables, they obtained the probability associated with different values of R_L. Their results, presented in the form of tables and a diagram, show that:

(1) The distributions of R_L are asymmetrical.

(2) The distributions of $\log R_L$ are more nearly symmetrical, but since R_L can only assume a finite number of discrete values, the distributions cannot in either form be represented adequately by smooth curves.

(3) Calculations from the partially grouped values give the results in Table I.

TABLE I

ρ	0·150	0·250	0·500	1·500
Mean (R_L)	0·164	0·284	0·558	1·648
$\sigma(R_L)$	0·066	0·117	0·226	0·689
$\log \rho$	−0·824	−0·602	−0·301	+0·176
Mean ($\log R_L$)	−0·816	−0·578	−0·285	+0·184
$\sigma(\log R_L)$	0·163	0·164	0·163	0·168
$\sigma(\log R_L)$ (limiting formula)	0·1535	0·1532	0·1768	0·1550

N.B. Logarithms are to base 10.

(4) The values of $\sigma(\log R_L)$ calculated from Halvorson & Ziegler's probability distribution remain nearly constant in the range of ρ considered. Having regard to the respective standard deviations, the bias of $\log R_L$ is of less importance than the bias of R_L. However, if a method of obtaining accurate fiducial limits were available, the bias in the single-valued estimate would be of no importance.

(5) The figures in the last row have been calculated from Fisher's formula for the large sample value of the variance of a maximum likelihood estimate, which reduces in this case to (Fisher, 1922, p. 364):

$$\{\sigma^2(\log_e R_L)\}^{-1} = 10\left\{\frac{e^{-10\rho}}{1-e^{-10\rho}}\,100\rho^2 + \frac{e^{-\rho}}{1-e^{-\rho}}\,\rho^2 + \frac{e^{-\frac{1}{10}\rho}}{1-e^{-\frac{1}{10}\rho}}\,\frac{\rho^2}{100}\right\}.$$

This formula gives minima for $\sigma(\log R_L)$ in the neighbourhood of $\rho = 0{\cdot}16$, $1{\cdot}6$ and 16. A series of calculated values including those tabled above are:

TABLE II

ρ	0·10	0·15	0·25	0·40	0·50	1·00	1·50	2·00	2·50	3·00	4·00
$\sigma(\log_{10} R_L)$	0·166	0·153	0·153	0·169	0·177	0·166	0·155	0·153	0·156	0·161	0·174

The probability values tabled by Halvorson & Ziegler are not sufficiently detailed to determine how far the differences between the values of $\sigma(\log R_L)$ compared in the last two rows of Table I are due to inadequacy in the large-sample variance formula. It is, however, clear that the standard error of the logarithm of the maximum likelihood estimate changes very little with ρ. Halvorson & Ziegler reached a similar result by noting that $\sigma(R_L)/\rho$ was very stable. The same result was noticeable in a series of calculations concerned with estimating the density of organisms in milk, made by Barkworth & Irwin (1938). These authors analysed the results of seven separate experiments in which 255 tubes were tested at each of four dilutions, namely 1 : 10, 1 : 50, 1 : 250, and 1 : 1250. They obtained the standard error of the maximum likelihood estimate of ρ from a large sample formula analogous to that given above; for seven values of ρ lying between 19 and 67, the ratio $\sigma(R_L)/\rho$ lay between 0·058 and 0·062.

It is clear that if more detailed computations of the probability distribution of R_L for a wider range of values of ρ were made on these lines, charts or tables giving fiducial or confidence limits for ρ could be readily supplied. In their paper of 1935 referred to above Matuszewski *et al.* have provided one such chart formed, as will be described below, on a basis which is partly empirical. The chart may be used as follows:

(*a*) Having observed experimentally the three numbers n_{10}, n_1, and $n_{0\cdot 1}$, obtain from Halvorson & Ziegler's tables the maximum likelihood estimate, R_L, of ρ.

(*b*) Taking R_L, read off from the chart lower and upper confidence limits, say $\rho_1(R_L)$ and $\rho_2(R_L)$.*

(*c*) Then, using Neyman's terminology, the statement

$$\rho_1(R_L) < \rho < \rho_2(R_L)$$

may be associated with a confidence coefficient of 0·95. In other words, if this procedure is applied in general bacteriological practice to the three frequencies n_{10}, n_1, and $n_{0\cdot 1}$ obtained from the dilution method, then the odds will be at

* The authors take λ' as the maximum likelihood estimate of mean density λ in the most concentrated of the three solutions, so that $\lambda' = 10R_L$.

least* 19 : 1 that the interval $\rho_1(R_L)$, $\rho_2(R_L)$ will cover the unknown mean density ρ.

It should be emphasized that the chart provided by Matuszewski *et al.* was not based on an exact mathematical solution, but found by graduating a series of experimental sampling results; the following quotation from their paper (p. 76) explains the method employed:

> The method followed by Miss J. Supińska consisted in a complex sampling experiment, using Tippett's random sampling numbers. The experiment produced a series of values of the variates x_0, x_1 and x_2 [i.e. n_{10}, n_1, and $n_{0\cdot 1}$ in Gordon's notation] following the sampling distribution which they would follow in our hypothetical conditions of the experiment. For each series of x_0, x_1 and x_2 it was possible to read up from the table of Halvorson & Ziegler an estimate, say λ', of the concentration λ. The estimates λ' have been then tabulated and an empirical frequency distribution of λ' corresponding to several fixed values of λ has been determined. Following the method described by J. Neyman, these empirical frequency distributions were then used to construct confidence intervals as if they were the accurate ones. As the random variation could not fail to affect the limits of the intervals it was felt necessary to correct them by fitting two parabolae, one marking the lower and the other the upper limits of the confidence intervals.

While therefore the writers did not claim exactness for their results, their chart, combined with Halvorson & Ziegler's table, does provide a provisional working solution not yet available to those who prefer the inverse approach.

It is interesting to obtain from the chart the limits $\rho_1(R_L)$ and $\rho_2(R_L)$ in the three hypothetical cases discussed by Gordon on p. 178 above. In Table III these limits and also their logarithms are given.

TABLE III

Observed frequencies			Gordon's estimate, $\bar{\rho}$	Maximum likelihood estimate, R_L	$\log R_L$	95 % confidence limits				
n_{10}	n_1	$n_{0\cdot 1}$				ρ_1	ρ_2	$\log \rho_1$	$\log \rho_2$	$\frac{1}{2}\log \rho_1\rho_2$
10	7	3	1·43	1·53	0·18	0·80	3·00	−0·097	0·477	0·19
8	5	1	0·291	0·267	−0·57	0·125	0·525	−0·903	−0·280	−0·59
4	2	1	0·086	0·080	−1·10	0·029	0·165	−1·54	−0·78	−1·16

It will be noted that:

(1) The 95 % confidence interval for ρ is in all cases relatively broad and, having regard to this, the differences between the single valued estimates $\bar{\rho}$ and R_L are of little importance.

(2) While neither $\bar{\rho}$ nor R_L is central with regard to the interval, log R_L

* 0·95 is a lower limit to the probability of a correct statement, owing to the discontinuous distribution of the frequencies.

differs only slightly from $\frac{1}{2}(\log \rho_1 + \log \rho_2)$, as will be seen by comparing the 6th and last columns. The length of the interval $\log \rho_1 - \log \rho_2$ changes slowly, increasing as R_L decreases. A study of the chart shows that for R_L roughly in the range 0·6–2·4, the breadth of the interval $\log \rho_2 - \log \rho_1$ remains nearly constant at a value of about 0·55, increasing slowly as R_L drops below 0·6 or rises above 2·4. Without further details of Supińska's sampling experiment, it is not possible to make a closer comparison between this confidence chart based on random sampling and the four distributions of R_L tabled by Halvorson & Ziegler. Both results suggest, however, that for a considerable range of values of ρ, $\log_{10} R_L$ is distributed about a mean value of approximately $\log_{10} \rho$ with a standard error of about 0·16.

If finally it be asked whether the direct or inverse solution is to be preferred, the answer must, I think, be that this can only be a matter of personal opinion. As mentioned in the footnote on p. 181 Greenwood & Yule (1917) preferred the latter method. The fundamental difference of the two methods of approach has recently been emphasized by Harold Jeffreys, with whom it is to be supposed that Gordon is in substantial agreement. Writing with regard to the *a priori* distribution of an unknown parameter, such as ρ, Jeffreys says (1938, p. 466):

> I can find nothing in the works of the pioneers of the principle of inverse probability to suggest that they identified the prior probability with a known frequency, and believe that if such an idea had occurred to them they would have repudiated it as definitely as I do. The function of a prior probability used to express ignorance is simply to express formally the transition from an inference about different possible data, given the hypothesis, to one about different hypotheses given the same data, and this transition must be made somehow on any theory.

It follows that if the prior probability distribution cannot be identified with frequency, neither can the posterior distribution. A probability, in Jeffreys' sense, obtained from the integral of the posterior distribution between limits ρ_1 and ρ_2, can be regarded as no more than a rational measure of the degree of belief that the experimenter may place in the truth of the statement $\rho_1 < \rho < \rho_2$. It can have no precise link with long-run relative frequency. This, indeed, it could only have if there were reason to suppose that in repeated dilution experiments the population value of ρ would be distributed uniformly between 0 and ∞.* But the legitimacy of any attempt to make this connexion Jeffreys has denied emphatically. For him, if I understand rightly, the posterior distribution derived from the formal prior distribution provides the one type of numerical scaling which he regards as useful in forming an opinion on the value of an unknown constant. To quarrel with this conviction would be out of place.

On the other hand, those who agree with Jeffreys must recognize that the direct approach with its fiducial argument has resulted from the development

* Gordon has assumed this to be the most appropriate form of prior distribution for ρ; possibly Jeffreys would prefer to make the distribution vary as $1/\rho$.

of a line of thought which finds probability statements useful in a form in which they can be directly related to long-run frequency of occurrence. To know that if a certain experimental technique is carried out and an arithmetical calculation made, then there are strong grounds for believing* that about 19 times out of 20 the limits $\rho_1(R_L)$ and $\rho_2(R_L)$ will include the unknown ρ is a form of information which appeals to a large number of statisticians. In this and other problems it seems likely to appeal particularly to persons who are frequently repeating the same form of operation, and can therefore the more readily appreciate the consequences of "being wrong" once in 10, once in 20, or once in 100 times, according to the risk they choose to allow. It is true that in certain instances there may be other special information which will enable the experimenter to guess at narrower or modified limits, rather than those obtained from the standard fiducial procedure. But whether this information is ever of a kind which can be put into numerical form is doubtful. Certainly this would not be achieved by using the prior probability distribution proportional to $d\rho$ or $d\rho/\rho$. The fact that we can sometimes narrow the range of uncertainty and so get nearer the mark, does not detract from the long-run "safeness" of the fiducial argument. Of course, whatever the approach, it is essential that the sampling should have been random.

From the point of view of the bacteriologist this difference of opinion between experts may be discouraging, but it has been shown more than once that the two lines of approach lead to results which, from the practical point of view, are almost precisely the same. Thus Jeffreys (1937), starting with a prior probability law for σ of $d\sigma/\sigma$, has reached "Student's" distribution for the posterior probability law for the mean; it follows that in this instance the fiducial limits of the direct approach associated with a confidence coefficient of, say, 0·95, will correspond exactly with those obtained from Jeffreys' posterior distribution and associated with this probability measure of 0·95. It is to be hoped, therefore, that if Mr Gordon obtains financial support for the lengthy computation required to produce tables of $\bar{\rho}$ and ϵ, he will at the same time make some research into the correspondence of his limits for ρ with those following from the direct approach. By this means he would undoubtedly widen the range of persons who could use his tables with confidence.

REFERENCES

BARKWORTH, H. & IRWIN, J. O. (1938). *J. Hyg., Camb.*, **38**, 446.

FISHER, R. A. (1922). *Philos. Trans.* A, **222**, 309.

GREENWOOD, M. & YULE, G. U. (1917). *J. Hyg., Camb.*, **16**, 36.

HALVORSON, H. O. & ZIEGLER, N. R. (1933*a*). *J. Bact.* **25**, 101.

—— —— (1933*b*). *J. Bact.* **26**, 331.

—— —— (1933*c*). *J. Bact.* **26**, 559.

JEFFREYS, H. (1937). *Proc. Roy. Soc.* A, **160**, 325.

—— (1938). *Proc. Roy. Soc.* A, **167**, 464.

MATUSZEWSKI, T., NEYMAN, J. & SUPIŃSKA, J. (1935). *J. R. Statist. Soc. Suppl.* **2**, 63.

* The confidence felt cannot be expressed in terms of a numerical probability but is based upon experience of the stability of statistical ratios.

ON THE DISTRIBUTION OF MAXIMUM LIKELIHOOD ESTIMATES

By B. L. WELCH

The properties of maximum likelihood estimates have been discussed by F. Y. Edgeworth (1908), R. A. Fisher (1922) and others. Most important is the property that in large samples a maximum likelihood estimate tends to be normally distributed with a variance which is given by a very simple formula and that no other estimate can have a smaller variance. If we have a random sample of n from a population whose probability law, $p(x \mid \theta)$, depends on only one parameter θ, and if T is the maximum likelihood estimate of θ, then under certain conditions it may be shown that T is in large samples normally distributed about θ with variance $1/nA_\theta$, where

$$A_\theta = \int \left(\frac{\partial \log p}{\partial \theta}\right)^2 p\,dx = -\int \frac{\partial^2 \log p}{\partial \theta^2} p\,dx. \qquad \text{......(1)}$$

Similar formulae are available for the variances and covariances of estimates when there are several parameters, but only the single parameter case will be discussed here.

When dealing with small samples the maximum likelihood estimate T is frequently adopted together with the method of approximating its distribution by referring it to a normal curve with mean θ and variance $1/nA_\theta$. The question arises whether this is an adequate procedure. This question splits into two parts. First, we may ask how far we are likely to go wrong by assuming that T is actually distributed in the manner known to be correct in large samples: and secondly, we may ask how far the advantage which T holds over any other estimate in the matter of sensitivity in large samples is retained in small samples. The first of these problems has an interesting historical aspect. Karl Pearson (1936) has told us that because of his doubt of the adequacy of the approximation in finite samples he made no subsequent use of the above large sample variance formula originally given by him and L. N. G. Filon (1898). However, the real reason for his view that the approximation is not good perhaps lay partly in the fact that he had not noted that the formula referred only to maximum likelihood estimates—a point not made clear until Edgeworth returned to the problem in 1908. Indeed this question of how good the approximation is cannot be answered very definitely, for there is, as a rule, no general agreement as to how close an approximation should be before it can be termed good.

The second problem has been discussed extensively by R. A. Fisher (e.g. 1925), who concludes that the maximum likelihood estimate still retains in

small samples desirable properties which recommend its use in preference to other estimates.

The present note is concerned only with the first problem and, in connexion with this, it is not intended to imply that the approximation by the large sample formula is not good but rather to give a method which may supply a closer approximation if this be thought desirable.

Let us write $p(x \mid \theta)$ equal to $\exp[\phi(x, \theta)]$ so that the maximum likelihood estimate T is given by

$$\Sigma \phi'(x, T) = 0. \qquad \ldots\ldots(2)^*$$

In general, the solution of this equation is not algebraically simple, although the numerical solution by iteration may not be difficult. It is therefore not possible to find the distribution of T, and indeed it is generally impracticable to find even the moments of T. However, usually the actual distribution of T is not required, but simply a method of calculating the probability that T shall exceed any specified value T_0 (say). It is often possible to find the moments of another quantity which will facilitate such a calculation.

If the sample be represented by a point in n-dimensional space, then all samples yielding the same value T_0 of T will lie on the hypersurface $\Sigma \phi'(x, T_0) = 0$. Conversely, with most probability laws likely to be encountered, and certainly for that of the example given below, it will not be possible for this hypersurface to contain points yielding a maximum likelihood estimate different from T_0. On one side of the hypersurface we may therefore expect $T > T_0$ and on the other side $T < T_0$. Now since by Taylor's theorem

$$\Sigma\phi'(x, T_0) = \Sigma\phi'(x, T) + (T_0 - T)\,\Sigma\phi''(x, T) + \text{etc.}, \qquad \ldots\ldots(3)$$

and since by definition of T we have $\Sigma\phi'(x, T) = 0$ and $\Sigma\phi''(x, T)$ negative, we shall expect $T > T_0$ on that side of the hypersurface for which $\Sigma\phi'(x, T_0) > 0$. Hence if the maximum likelihood equation establishes a many-one relationship between sample points and values of T, we see that the probability that $T > T_0$ is equal to the probability that $\Sigma\phi'(x, T_0) > 0$. But the quantity $\Sigma\phi'(x, T_0)$ is the simple sum of a number of independent components, and its cumulants therefore follow simply from those of a single $\phi'(x, T_0)$. Any of the usual methods of approximating to probabilities from cumulants may then be employed to evaluate $P\{\Sigma\phi'(x, T_0) > 0\}$ and hence $P(T > T_0)$.

As an example consider the probability law

$$p(x \mid \theta) = \frac{2}{\Gamma(\frac{1}{4})} e^{-(x-\theta)^4}, \qquad \ldots\ldots(4)$$

which has been discussed at some length by Edgeworth (1908). We have $\phi'(x, \theta) = 4(x - \theta)^3$ and therefore

$$\Sigma\phi'(x, T_0) = 4\Sigma(x - T_0)^3. \qquad \ldots\ldots(5)$$

* The dash denotes differentiation with respect to θ, and Σ denotes summation over the whole sample.

Now by a straightforward method it may be shown that the first three cumulants of $4(x-T_0)^3$, when the true value of the parameter is θ, are given by

$$\left.\begin{aligned} \kappa_1 &= \mu_1' = 12c(\theta-T_0)+4(\theta-T_0)^3, \\ \kappa_2 &= \mu_2 = 12c+(60-144c^2)(\theta-T_0)^2+144c(\theta-T_0)^4, \\ \kappa_3 &= \mu_3 = 36(\theta-T_0)[(5-12c^2)+(48c+96c^3)(\theta-T_0)^2 \\ &\qquad +(36-144c^2)(\theta-T_0)^4], \end{aligned}\right\} \quad \ldots\ldots(6)$$

where
$$c = \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})} = 0\cdot33799.$$

The cumulants of $\Sigma\phi'(x,T_0)/n$ are then κ_1, κ_2/n and κ_3/n^2. By fitting a Pearson Type III curve with these three moments we can then approximate to $P\{4\Sigma(x-T_0)^3/n>0\}$. If closer approximations are necessary further moments may be calculated and some other kind of curve fitted.

More frequently, perhaps, the converse problem is posed of finding T_0 such that $P(T>T_0)$ has a specified probability. The above method then involves a certain amount of trial and error. For instance, suppose $n=25$ and T_0 is wanted so that $P(T>T_0)$ equals 0·05. A first approximation is obtained by taking T to be normally distributed about θ with variance $1/25A_\theta$, where

$$A_\theta = \int\{4(x-\theta)^3\}^2 p\,dx = 12\frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})} = 4\cdot056. \quad \ldots\ldots(7)$$

The approximate standard deviation of T is then 0·0993 and T_0 is $\theta+0\cdot163$. Substituting this value of T_0 in (6) we obtain $\kappa_1=-0\cdot678$, $\kappa_2/25=0\cdot210$, and $\kappa_3/625=-0\cdot0392$. The standard deviation of $4\Sigma(x-\theta-0\cdot163)^3/25$ is therefore exactly 0·458, and the skewness $\sqrt{\beta_1}=-0\cdot407$. Using these moments and the very convenient tables of L. R. Salvosa (1930) we find that approximation by a Pearson Type III curve gives $P\{4\Sigma(x-\theta-0\cdot163)^3/25>0\}=0\cdot056$. From this it is not difficult to see that, by adjusting T_0 to the value $\theta+0\cdot168$, the Type III approximation to $\Sigma\phi'(x,T_0)/n$ will give $P(4\Sigma(x-\theta-0\cdot168)^3/25>0\}=0\cdot05$, i.e. $P(T>\theta+0\cdot168)=0\cdot05$.

There is in the present example little difference between the results given by the large sample approximation to the distribution of T and the Type III approximation to the distribution of $\Sigma\phi'(x,T_0)/n$. This is partly due to the nature of $p(x\mid\theta)$, owing to whose symmetry the maximum likelihood estimate, T, has a distribution which is *exactly* symmetrical with true mean at θ. The main approximation in using the large sample method therefore lies in the value adopted for the variance. Often the distribution of T will not be symmetrical nor have true mean exactly at θ. It is in such cases that a method which uses exact values of the lower order moments of $\Sigma\phi'(x,T_0)/n$ may be expected to yield the greatest improvement in accuracy.

REFERENCES

Edgeworth, F. Y. (1908, 1909). "On the probable errors of frequency constants." *J. R. Statist. Soc.* 71, 381–97, 499–512, 651–78; **72**, 81–90.

Fisher, R. A. (1922). "On the mathematical foundations of theoretical statistics." *Philos. Trans.* A, **222**, 309–68.

Fisher, R. A. (1925). "Theory of statistical estimation." *Proc. Camb. Phil. Soc.* **22**, 700–25.

Pearson, Karl (1936). "Method of moments and method of maximum likelihood." *Biometrika*, **28**, 34–59.

Pearson, Karl & Filon, L. N. G. (1898). *Philos. Trans.* A, **191**, 229 *et seq.*

Salvosa, L. R. (1930). *Tables of Pearson's Type III Function.* Reprinted from *Ann. Math. Stat.* May 1930. Publ. Edwards Brothers, Inc., Ann Arbor, Michigan.

ON NEYMAN'S "SMOOTH" TEST FOR GOODNESS OF FIT

I. DISTRIBUTION OF THE CRITERION ψ^2 WHEN THE HYPOTHESIS TESTED IS TRUE

By F. N. DAVID

1. Introductory

The χ^2 test has passed into common use since its introduction by Karl Pearson in 1900. It has, however, long been recognized that in certain cases it is inadequate as a test for goodness of fit, and in particular in the case when the deviations of the observations from the hypothesis tested are consecutively positive (or negative); for the χ^2 test, by taking into account the square of the difference between the observed and expected values, renders it impossible to pay attention to the sign of this difference. It is because of this inadequacy that Neyman (1937) introduced the "smooth" test for goodness of fit. This "smooth" test has been discussed by E. S. Pearson (1938) in some detail, which absolves us from further discussion here, but the procedure of the test will be set out in order that the substance of this present paper may be understood.

Following the procedure of previous papers (1933, 1936) Neyman (1937, p. 156) insists that a choice of a suitable test for goodness of fit may only be made after a statement of the probability functions specified both by the hypothesis tested and by the admissible hypotheses alternative to that tested. In order to supplement the χ^2 test he confines the alternative frequency laws to a class of functions which he terms "smooth", and whose form it is supposed can be represented by transformed Legendre polynomials. The keystone of his test lies in a probability integral transformation previously given by R. A. Fisher (1932) and K. Pearson (1933), by means of which the distribution of n independent random variables may be made rectangular.*

The criterion ψ^2, which may be used instead of χ^2 if the admissible probability laws are smooth, is designed to test the departure of the distribution of the transformed variables from rectangularity. Before it may be applied, however, it is necessary to choose what may be termed the order of the test, bearing in mind which departures from the hypothesis tested it is most important for the investigator not to overlook. The question of the appropriate order of test to choose has

* For example if

$$\left.\begin{aligned} p(y) &= \text{constant for } a<y<b \\ &= 0 \text{ otherwise} \end{aligned}\right\}$$

we should say that y was distributed rectangularly in the interval $[a; b]$. An alternative method of expression would be to say that in the interval $[a; b]$ all values of y are equally likely.

been discussed to some extent by Neyman (1937) and E. S. Pearson (1938). Further investigation into the matter has been carried out and the results will be published in a second part of this paper. Once the order of the test has been decided, then the calculation of the criterion ψ^2 is a comparatively simple matter and may be set out in the following way.

Assume that there are n independent random variables x_j $(j=1, 2, \ldots, n)$, following a known continuous probability law which is specified by H_0, the hypothesis to be tested. By means of the relation

$$y=\int_{\infty}^{x} p(x/H_0)\,dx, \qquad \ldots\ldots(1)$$

the n independent variables x are transformed into n independent variables y which are rectangularly distributed, *whatever the probability law of the x's*. If it is decided that a test of the kth order is to be applied, the next step involves the calculation of k quantities u_i $(i=1, 2, \ldots, k)$ by means of the substitution of the n variables y which were obtained from (1) into the appropriate Legendre polynomials, π_i, where

$$u_i=n^{-\frac{1}{2}}\sum_{j=1}^{n}\pi_i(y_j) \quad i=1, 2, \ldots, k. \qquad \ldots\ldots(2)$$

The first four Legendre polynomials* are

$$\left.\begin{array}{ll}\pi_1(y)=\sqrt{12}\,(y-\frac{1}{2}) & \pi_3(y)=\sqrt{7}\,(20(y-\frac{1}{2})^3-3(y-\frac{1}{2}))\\ \pi_2(y)=\sqrt{5}\,(6(y-\frac{1}{2})^2-\frac{1}{2}) & \pi_4(y)=210(y-\frac{1}{2})^4-45(y-\frac{1}{2})^2+\frac{9}{8}\end{array}\right\} \qquad \ldots\ldots(3)$$

The criterion ψ^2 consists of the sum of the squares of these u's, thus

$$\psi^2=\sum_{i=1}^{k}u_i^2=n^{-1}\sum_{i=1}^{k}\left(\sum_{j=1}^{n}\pi_i(y_j)\right)^2. \qquad \ldots\ldots(4)$$

Any such criterion is, however, of little use if its distribution is not known. An approximation to the distribution was given by Neyman and it is this approximation which is reviewed in the next section.

2. Distribution of ψ^2

It was shown by Neyman that when n, the number of observations in the sample, is large, then each of the quantities u_i^2 $(i=1, 2, \ldots, k)$, is distributed independently as χ^2 with one degree of freedom. It will follow therefore that for large n the criterion ψ^2 of a kth order test will be distributed as χ^2 with k degrees of freedom. As with many statistical tests based on the assumption that n is large, the decision as to the numerical value of n is left to the user of the test and, with the introduction of the personal element, we get a variation in ideas both as to

* It is of interest to note that a recurrence formula for these polynomials follows directly from a general formula for Legendre polynomials (see for example Whittaker & Watson, *Modern Analysis*, p. 302, II). If $z=y-\frac{1}{2}$ then

$$\frac{n+1}{\sqrt{(2n+3)}}\,\pi_{n+1}(z)-2z\sqrt{(2n+1)}\,\pi_n(z)+\frac{n}{\sqrt{(2n-1)}}\,\pi_{n-1}(z)=0.$$

what is meant by a large sample and the point at which a large sample becomes a small sample. It is therefore useful to obtain some idea of the degree of approximation involved in the supposition that ψ^2 is distributed as χ^2. Although this may only be investigated completely by the discovery of the true distribution of ψ^2 it is not difficult to find Pearson curves which have moments agreeing with the true moments of ψ^2, and these may be used to throw light on the approximation to the χ^2 distribution.

The derivation of the true moments of ψ^2 is a matter of simple but tedious algebra. Using the notation $\mathscr{E}$ to denote mathematical expectation, it is seen that we must calculate

$$\mu_t = \mathscr{E}(\psi^2 - \mathscr{E}(\psi^2))^t \quad \text{for} \quad t = 1, 2, 3, 4, \qquad \ldots\ldots(5)$$

where ψ^2 is as defined in (4) and the y's are independently distributed within the interval [0; 1] in accordance with the rectangular law. I have calculated only the moments of ψ^2 for first and second order tests, partly because numerical work* leads me to believe that these order tests are more powerful to detect *any* departures from the basic hypothesis H_0 than any other higher order test, and partly because the necessary algebra became so very heavy.

From the properties of the transformed Legendre polynomials we know (see Neyman, 1937) that

$$\mathscr{E}(\pi_i(y_j)) = \int_0^1 \pi_i(y_j)\,dy_j = 0 \quad \text{for any } i \text{ and } j. \qquad \ldots\ldots(6)$$

Also since the observations are independent of one another

$$\mathscr{E}(\pi_i(y_j)\,.\,\pi_i(y_p)) = 0, \qquad \ldots\ldots(7)$$

and

$$\mathscr{E}(\pi_i^2(y_j)) = \int_0^1 \pi_i^2(y_j)\,dy_j = 1. \qquad \ldots\ldots(8)$$

In general it will be noticed that we require to calculate

$$\mathscr{E}(u_i^2)^t = n^{-1}\mathscr{E}\left(\sum_{j=1}^{n} \pi_i(y_j)\right)^{2t} \quad \text{for } i = 1,\, 2 \text{ and } t = 1,\, 2,\, 3,\, 4 \qquad \ldots\ldots(9)$$

and

$$\mathscr{E}(u_i^{2p}\,u_m^{2q}) = n^{-1}\mathscr{E}\left(\left(\sum_{j=1}^{n} \pi_i(y_j)\right)^{2p}\left(\sum_{j=1}^{n} \pi_m(y_j)\right)^{2q}\right) \quad \text{for } i, m = 1, 2 \text{ and } p, q = 1, 2, 3. \qquad \ldots\ldots(10)$$

The calculations required are straightforward and follow at once from the application of the elementary theorems concerning the sum and product of expectations. The final results are given below in Table I.

It will be noticed that for n small the moments of ψ_1^2 and ψ_2^2 differ considerably from those of χ^2. It is, however, only fair to point out here that Neyman supposes, as in the case of the χ^2 test, that a "smooth" test for goodness of fit would not be

* To be discussed in Part II of this paper.

TABLE I

True moments of ψ^2	$\psi_1^2 = u_1^2$, $(k=1)$	$\psi_2^2 = u_1^2 + u_2^2$, $(k=2)$
μ_1'	1	2
μ_2	$2-\frac{6}{5n}$	$4-\frac{32}{35n}$
μ_3	$8-\frac{72}{5n}+\frac{48}{7n^2}$	$16+\frac{704}{49n}-\frac{722208}{35035n^2}$
μ_4	$60-\frac{936}{5n}+\frac{7524}{35n^2}-\frac{432}{5n^3}$	$144+\frac{15216}{49n}-\frac{2203468}{35035n^2}+\frac{17946980}{119119n^3}$

carried out except on a large number of observations. It seems to be possible by an approximate method to obtain some idea of how large n should be.

A consideration of the β-coefficients calculated from the moments of Table I suggests that the most appropriate curves to use would be a Pearson Type I for ψ_1^2 and Type VI for ψ_2^2. Since, however, it was hoped that both curves would approximate quickly to the χ^2 Type III distribution, as a first step Type III curves with the start of the curve at the origin were fitted to the moments of both ψ_1^2 and ψ_2^2. As there are only two constants to evaluate for the Type III curve with zero start, the procedure was equivalent to finding a χ^2 distribution with the mean and standard deviation equal to the corresponding true values for ψ^2. For purposes of comparison, the 5 % and 1 % levels of χ^2 with $f = 1$ and $f = 2$ were taken, and the tail areas of the fitted Type III curve corresponding to these abscissae were calculated from *Tables of the Incomplete Γ-Function* (K. Pearson, 1922). The results are summarized in Table II.

TABLE II

$f=1$, $\chi^2_{0\cdot05}=3\cdot841$, $\chi^2_{0\cdot01}=6\cdot635$						
n	5	10	20	30	50	100
μ_2	1·76	1·88	1·94	1·96	1·976	1·988
$P\{\psi_1^2 > 3\cdot841\}$	0·0523	0·0511	0·0505+	0·0504	0·0502+	0·0502−
$P\{\psi_1^2 > 6\cdot635\}$	0·0098	0·0099	0·0099[5]	0·0100	0·0100	0·0100

$f=2$, $\chi^2_{0\cdot05}=5\cdot991$, $\chi^2_{0\cdot01}=9\cdot210$						
n	5	10	20	30	50	100
μ_2	3·8171	3·9086	3·9543	3·9695+	3·9817	3·9909
$P\{\psi_2^2 > 5\cdot991\}$	0·0511	0·0505+	0·0503	0·0502	0·0501	0·0500
$P\{\psi_2^2 > 9\cdot210\}$	0·0100	0·0100	0·0100	0·0100	0·0100	0·0100

The probability levels of ψ^2 as given by the empirical Type III curve approximate quite quickly to those of χ^2, and at first sight it would appear that for any size of sample little risk would be involved in testing the significance of ψ^2 by means of the χ^2 tables. However, it must not be forgotten that the empirical curve is itself an approximation to the unknown true ψ^2 distribution. To obtain some idea therefore of the approximation involved it is necessary to calculate the β_1 and β_2 of the fitted Type III curves and compare them with the β's of the true distribution of ψ^2 and of the χ^2 distribution. These values of the β's are plotted in the diagram and their numerical values are given in Table IV. It will be noted that the β's of the fitted Type III curves lie nearer the χ^2 points (8, 15) and (4, 9) than do those of the true ψ^2. Hence, while the results of Table II suggest that the approximation of the ψ^2 distribution by a Type III distribution is a fairly good one, it would appear that further investigation is advisable before any definite conclusion may be drawn.

It was therefore decided to utilize the true third moments of ψ^2. Type I curves with zero start were fitted using the first three moments of ψ_1^2, and similarly Type VI curves using those of ψ_2^2. It will be recognized that the procedure for the fitting of a Type VI curve and the evaluation of its integral is very similar to that for the fitting of Type I, and the calculations required were almost the same for both curves. Accordingly it is necessary to set out the steps followed for the case of Type I only.

If the equation of the Type I curve is written as

$$Z = \frac{1}{a^{m_1+1}} \frac{\Gamma(m_1+m_2+2)}{\Gamma(m_1+1)\,\Gamma(m_2+1)}\, Y^{m_1}\left(1-\frac{Y}{a}\right)^{m_2} \qquad \text{......(11)}$$

easy algebra will give

$$M_1' = a\frac{(m_1+1)}{(m_1+m_2+2)}, \quad M_2' = aM_1'\frac{(m_1+2)}{(m_1+m_2+3)}, \quad M_3' = aM_2'\frac{(m_1+3)}{(m_1+m_2+4)}, \qquad \text{......(12)}$$

where M_1', M_2' and M_3' are the true moments of ψ_1^2 about the origin. It was not found possible to write down the solutions of a, m_1 and m_2 in any neat form, and the procedure followed was to substitute numerical values for M_1', M_2' and M_3' at an early stage. The partial integral of (11) is recognized as an Incomplete Beta-Function. Hence to calculate the tail areas corresponding to $\chi^2_{0\cdot05}$ and $\chi^2_{0\cdot01}$ it will only be necessary to refer to *Tables of the Incomplete Beta-Function* (K. Pearson, 1934). Table III gives the results of interpolation into these tables.

It will be noted that for strict accuracy triple interpolation would be necessary. This, however, called for much calculation, so an approximation was made by taking $m_1 = -0\cdot50$ and using a double interpolation formula only. The effect of the approximation is to make the interpolated values slightly *greater** numerically than those values which would have been obtained by the use of triple inter-

* A rough form of quadrature for the case $f=1$, $n=5$ gives $P\{\psi_1^2 > 3\cdot841\} = 0\cdot0481$.

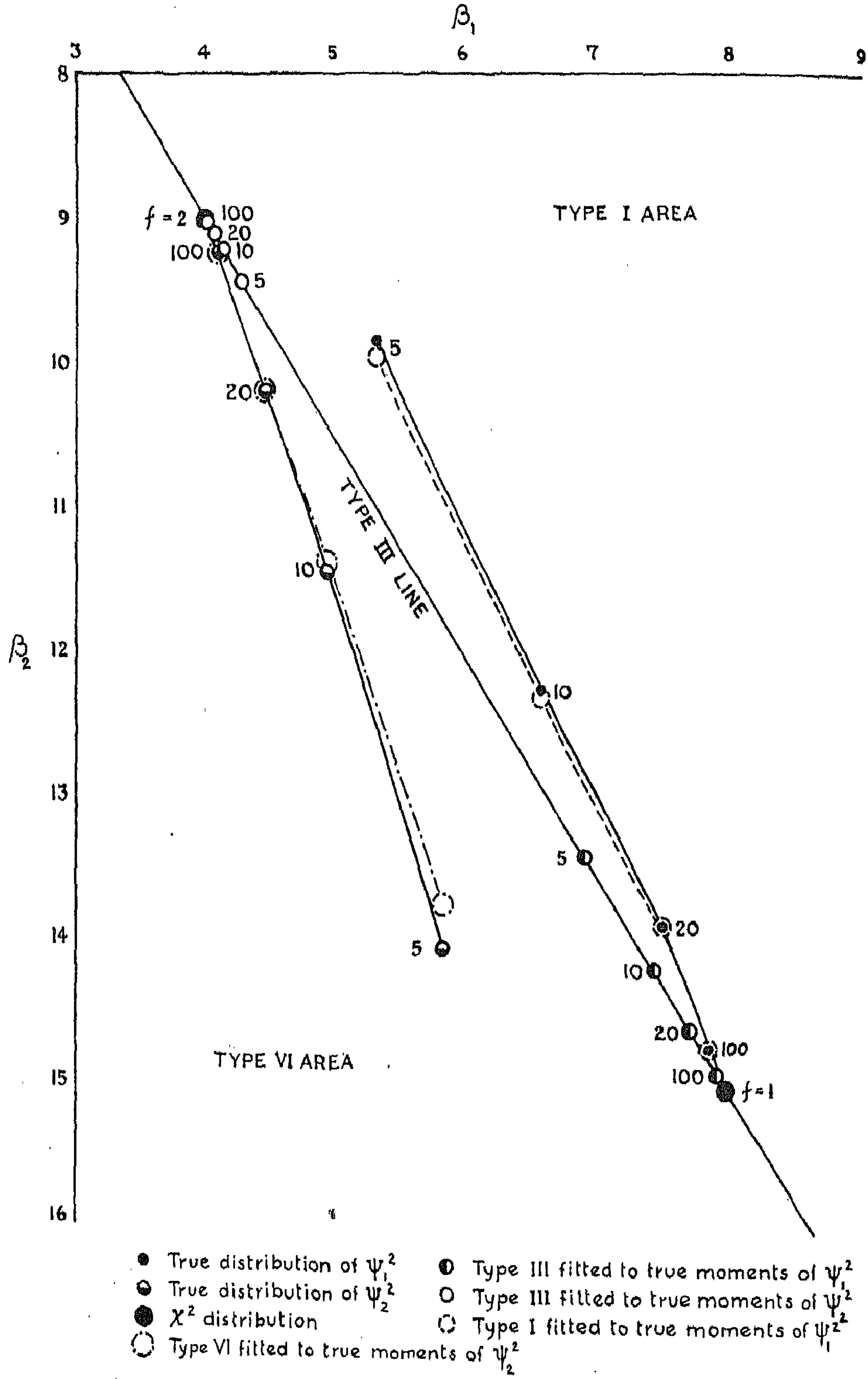


Fig. 1. β_1, β_2 points for true and approximate distributions.

TABLE III

$f=1$, $\chi^2_{0\cdot05}=3\cdot841$, $\chi^2_{0\cdot01}=6\cdot635$				$f=2$, $\chi^2_{0\cdot05}=5\cdot991$, $\chi^2_{0\cdot01}=9\cdot210$			
n	5	10	20	n	5	10	20
μ_3	5·3944	6·6286	7·2971	μ_3	18·0489	17·2312	16·6668
a	20·5920	45·3643	95·2355$^+$	a	16·2813	30·8654	60·3289
m_1	−0·50797	−0·50185$^+$	−0·50044	m_1	0·29947	0·154503	0·07825$^-$
m_2	8·63985$^+$	21·10009	46·07631	m_2	10·57853	17·81708	33·52469
$P\{\psi_1^2>3\cdot841\}$	0·049$^-$	0·049$^+$	0·050$^-$	$P\{\psi_2^2>5\cdot991\}$	0·047$^+$	0·049$^+$	0·050$^-$
$P\{\psi_1^2>6\cdot635\}$	0·007$^+$	0·009$^-$	0·010$^-$	$P\{\psi_2^2>9\cdot210\}$	0·010$^-$	0·010$^-$	0·010$^-$

TABLE IV

	Type of curve ...	...	$n=5$	$n=10$	$n=20$	$n=100$
ψ_1^2	True values	β_1 β_2	5·34 9·84	6·61 12·25$^+$	7·52 13·86^5	7·86 14·71
	Fitted Type III	β_1 β_2	6·95 13·42	7·48 14·22	7·74 14·61	7·95$^-$ 14·92
	Fitted Type I	β_1 β_2	5·34 9·94	6·61 12·30	7·52 13·89	7·86 14·71
ψ_2^2	True values	β_1 β_2	5·86 14·06	4·97 11·43	4·49 10·19	4·10 9·24
	Fitted Type III	β_1 β_2	4·30 9·45	4·14 9·21	4·08 9·12	4·01 9·01
	Fitted Type VI	β_1 β_2	5·86 13·74	4·97 11·37	4·49 10·17	4·10 9·23

polation formulae. As n increases m_1 gradually approaches the value $-0\cdot5$. For values of n greater than 20 the required Beta-Function ratio fell outside the range of the existing tables. These values therefore could not be evaluated except by quadrature but the results for $n = 5$, 10 and 20 give sufficient indication for our purposes of the approximate error involved. A system of triple interpolation was carried out for the evaluation of the integral of the Type VI curve but the calculations became so very heavy that it was only found practicable to obtain the result correct to three decimal places. The results of interpolation were confirmed approximately by a rough quadrature of the curves. The tail areas of the Type I

and Type VI curves beyond $\chi^2_{0\cdot05}$ are seen to be slightly less than 0·05 but nearer to that value than the corresponding areas from the Type III approximation shown in Table II. On the other hand the Type I and Type VI areas beyond $\chi^2_{0\cdot01}$ are a little further from 0·01 than for the Type III approximation. Except perhaps in the case of $n = 5$ of the Type I curve these differences can hardly be regarded as of importance. To judge how closely the Type I and Type VI curves are likely to approximate to the unknown true distributions we may compare their β_2 values with the true values. This has been done in Table IV and in the diagram. The proximity in position between the solid and dotted circles in the latter, suggests that the Type I and Type VI curves must represent the true ψ_1^2 and ψ_2^2 distributions very closely.

Thus the probability levels of the Type I and Type VI curves given in Table III may be taken as lying not very far from the true values, and this shows that no great error will be made if we assume for samples of 20 and over that Neyman's ψ^2 criterion of the first and second orders are distributed as χ^2 with one and two degrees of freedom respectively.

3. Sphere of usefulness of the "smooth" test

It was pointed out in § 1 that where the deviations of the observations from the hypothesis tested are consecutively positive (or negative) then the ψ^2 "smooth" test for goodness of fit may be applied. It does not seem possible to give any definite rule as to when it might be used in preference to the χ^2 test. Neyman remarks that it would be interesting to enquire into the relative workings of the χ^2 and ψ^2 tests, possibly with this point in mind, but nothing has been written on the subject and it is difficult to see how a comparison of the powers of the two tests could be obtained.

While recognizing the ingenuity of Neyman's "smooth" test, and the attempt it makes usefully to supplement the χ^2 test, the present writer feels that as it stands it is not wholly satisfactory, and that much work remains to be done before it can pass into common use. It has been pointed out in § 1 that Neyman (1937) and E. S. Pearson (1938) have both discussed the appropriate order of test to choose, without however reaching any definite conclusion. It has been possible to remedy this in part by confirming E. S. Pearson's suggestions through numerical calculations, but an extension of the theory appears necessary before the test is generally applicable.

In § 2 it has been shown that serious error will not be made if the use of the "smooth" test is restricted to samples of twenty and over, and Neyman himself says that he would use the test only if the available sample is large. At present, however, the application of the test presents rather a formidable task for any computer. For example, if the sample is of size 100, and it is desired to apply a test of the second order, the numerical work would entail 100 entries into tables of the appropriate probability integral with possible interpolations and 100

substitutions into each of the first and second polynomials. Hence for a sample greater in size than 100 the labour of computation involved will incline the computer towards the simpler and more familiar χ^2 test. The need for such a "smooth" test cannot be denied, but it would have greater utility if it could be applied after some grouping of the observations has been made. Whether this is a possible development or not I do not know.

At present it is only possible to apply the "smooth" test when the parameters of the hypothesis tested are known, whereas the χ^2 test permits one to calculate them from the data, as for example in the fitting of Pearson curves. What the effect of calculating parameters from the data would be on the ψ^2 criterion I cannot at present suggest, although it is hoped to throw light on this point by means of a sampling experiment. Certainly if parameters are calculated from the data and used in the basic hypothesis, H_0, it would mean that the variables y would be no longer independent or rectangularly distributed. Whether it is possible to determine what form the probability law of y would take, given restrictions on the original sample, remains to be investigated.

I have to thank Prof. E. S. Pearson for helpful criticism and Miss J. Townend for drawing the figure.

REFERENCES

FISHER, R. A. (1932). *Statistical Methods for Research Workers*, Fourth Edition, § 21·1.
NEYMAN, J. (1937). *Skand. Aktuar. Tidskr.* **20**, 149–99.
NEYMAN, J. & PEARSON, E. S. (1933). *Philos. Trans.* A, **231**, 289–337.
—— —— (1936). *Statist. Res. Mem.* **1**, 1–37.
PEARSON, E. S. (1938). *Biometrika*, **30**, 134–48.
PEARSON, K. (1933). *Biometrika*, **25**, 379–410.
—— (1922). *Tables of the Incomplete Γ-Function*. Biometrika Office.
—— (1934). *Tables of the Incomplete* B-*Function*. Biometrika Office.

TESTS OF HYPOTHESES CONCERNING LOCATION AND SCALE PARAMETERS

By E. J. G. PITMAN
University of Tasmania

1. The comparison of location parameters

Suppose that k observed numbers

$$x_1, x_2, \ldots, x_k,$$

are values of k chance variables, and that the elementary probability function of the simultaneous distribution of the chance variables is

$$F(x_1 - a_1, \ldots, x_k - a_k),$$

the function F being of known form but the values of the location parameters $a_1, a_2, \ldots, a_k$ being unknown. Suppose, further, that we wish to test the hypothesis that

$$a_1 = a_2 = \ldots = a_k,$$

which we shall call the hypothesis H_0. Any test must be based upon some statistic J which is a function of the observed values $x_1, x_2, \ldots, x_k$, and H_0 will be accepted for certain values of J and rejected for all other values. If the observations are our only source of knowledge of the values of the location parameters, a satisfactory test must give the same answer when the observed values are

$$x_1 + \lambda, x_2 + \lambda, \ldots, x_k + \lambda,$$

as when they are $\qquad x_1, x_2, \ldots, x_k.$

Hence the statistic J must have the property

$$J(x_1 + \lambda, \ldots, x_k + \lambda) \equiv J(x_1, \ldots, x_k). \qquad \ldots\ldots(1)$$

Without loss of generality, we may assume that J is always positive, and that a small value of J is regarded as significant, i.e. as indicating that the hypothesis H_0 is untrue. If this is not so, we simply replace J by some suitable function of it. In order to use the observed value of J as a test of H_0, we must know the distribution of J when H_0 is true. Since J has the property (1), its value will be unaffected by the same change of origin of all the x. Hence, when the a are all equal, we may, without loss of generality, assume that their common value is 0. We require then the distribution of J when

$$a_1 = a_2 = \ldots = a_k = 0.$$

Let p be any given number (such as 0·95) between 0 and 1. When this distribution of J is known, we can determine θ such that, when H_0 is true, the probability that $J \geqq \theta$ is p. Having chosen p, we reject the hypothesis H_0 if the observed value of

J is less than θ, and accept it otherwise. The probability of rejecting H_0 when it is true will be $1-p$. If J is so chosen that for a fixed value of θ the probability that $J \geqq \theta$ is greatest when H_0 is true, the hypothesis H_0 will be more likely to be accepted when it is true than when any alternative hypothesis is true. A test which has this property is said to be "unbiased" (Neyman & Pearson, 1936, p. 8). Let us attempt to determine J so that the resulting test is unbiased.

A set of values of the x may be specified by a point (the sample point) whose rectangular co-ordinates in a k dimensional space are $(x_1, x_2, \ldots, x_k)$. For the co-ordinates of a variable point in this space we shall use $(\xi_1, \xi_2, \ldots, \xi_k)$. Since J has the property (1), it will be constant along any line parallel to the line

$$\xi_1 = \xi_2 = \ldots = \xi_k, \qquad \ldots\ldots(2)$$

and therefore the region A, consisting of all points for which $J \geqq \theta$, will be bounded by a cylindrical hypersurface with its generators parallel to the line (2). When H_0 is true, the probability that $J \geqq \theta$ is

$$p = \int_A F(\xi_1, \ldots, \xi_k)\, d\xi_1 \ldots d\xi_k,$$

and when H_0 is not true, the probability is

$$\begin{aligned} p' &= \int_A F(\xi_1 - a_1, \ldots, \xi_k - a_k)\, d\xi_1 \ldots d\xi_k \\ &= \int_B F(\xi_1, \ldots, \xi_k)\, d\xi_1 \ldots d\xi_k, \end{aligned}$$

where B is a region formed from A by translation without rotation. Thus, the necessary and sufficient condition for an unbiased test is that the integral of $F(\xi_1, \ldots, \xi_k)$ over the region A be greater than its integral over any equal and parallel region B.

If L is any line parallel to the line (2), we shall write

$$P(L) = \int_L F(\xi_1, \ldots, \xi_k)\, d\eta,$$

where
$$\eta = \Sigma\xi/\sqrt{k},$$

the distance of the point $(\xi_1, \xi_2, \ldots, \xi_k)$ from the plane $\Sigma\xi = 0$, and the integral is taken along L. Now if A is defined as the locus of all lines L for which $P(L)$ is greater than or equal to some constant h, $P(L)$ will be less than h on any line L in B which is not also in A. From this it easily follows that

$$\int_A F(\xi_1, \ldots, \xi_k)\, d\xi_1 \ldots d\xi_k > \int_B F(\xi_1, \ldots, \xi_k)\, d\xi_1 \ldots d\xi_k,$$

and so the resulting test is unbiased. Hence we shall have an unbiased test if we define J at any point as equal to $P(L)$ for the line L through that point.

The co-ordinates of a variable point on the line L through the point

$$(x_1, x_2, \ldots, x_k)$$

may be expressed in the form

$$\xi_r = x_r - t \quad (r = 1, 2, \ldots, k)$$

where
$$t = \frac{\Sigma x}{k} - \frac{\eta}{\sqrt{k}}.$$

The expression for $P(L)$ becomes

$$P(L) = \sqrt{k}\int_{-\infty}^{\infty} F(x_1 - t, \ldots, x_k - t)\, dt.$$

Instead of $P(L)$ we shall take $P(L)/\sqrt{k}$ for J, and, replacing the symbol t by the symbol a, we have

$$J = \int_{-\infty}^{\infty} F(x_1 - a, \ldots, x_k - a)\, da.$$

The test based on this J will be unbiased.

As a simple illustration, take the case where the x are independent normal variables, each with unit standard deviation, and the mean value of x_r is a_r.

$$F(x_1 - a_1, \ldots, x_k - a_k) = \frac{1}{(2\pi)^{\frac{1}{2}k}} \exp\{-\tfrac{1}{2}\Sigma(x_r - a_r)^2\},$$

therefore
$$J = \frac{1}{(2\pi)^{\frac{1}{2}k}} \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}\Sigma(x-a)^2\}\, da$$
$$= \frac{e^{-\frac{1}{2}S}}{k^{\frac{1}{2}}(2\pi)^{\frac{1}{2}(k-1)}},$$

where
$$S = \Sigma(x_r - \bar{x})^2, \quad \bar{x} = \Sigma x_r/k.$$

In practice we would take S, which is a monotonic function of J, as our criterion, large values of S being significant. When H_0 is true, S is distributed like χ^2 with $k-1$ degrees of freedom.

Suppose now that the hypothesis to be tested asserts not merely that the location parameters of the chance variables x are all equal, but that these location parameters all have the same particular given value. By a change of origin of all the x this particular value may be taken as 0. Thus the hypothesis to be tested is that

$$a_1 = a_2 = \ldots = a_k = 0,$$

which we shall call the hypothesis H_0'. It is obvious that if we take

$$J' = F(x_1, \ldots, x_k)$$

and regard small values of J' as significant, the test will be unbiased. Moreover, it is very easy to show that, if the x are independent, so that F is the product of the separate elementary probability functions, and if each elementary probability function is unimodal, then, when H_0' is not true, the probability that $J' \geqq \theta$ increases as any a moves towards the value 0.

2. THE COMPARISON OF SCALE PARAMETERS

Here the elementary probability function of the chance variables x is

$$\Pi\{1/c_r\}\, F\{x_1/c_1, \ldots, x_k/c_k\},$$

where the c are all positive and we wish to test the hypothesis,

$$c_1 = c_2 = \ldots = c_k,$$

which we shall call the hypothesis H_1. Let us assume for the moment that the x are all positive chance variables. Putting

$$y_r = \log x_r, \quad a_r = \log c_r,$$

we have for the elementary probability function of the chance variables y

$$\exp(\Sigma y_r - \Sigma a_r)\,.\, F\{\exp(y_1 - a_1), \ldots, \exp(y_k - a_k)\}.$$

The hypothesis H_1 applied to the parameters c is equivalent to H_0 applied to the parameters a. We shall therefore have an unbiased test if we take as our criterion

$$J = \int_{-\infty}^{\infty} \exp(\Sigma y_r - ka)\,.\, F\{\exp(y_1 - a), \ldots, \exp(y_k - a)\}\, da$$

$$= \Pi\{x_r\} \int_0^{\infty} \frac{F\{x_1/c, \ldots, x_k/c\}}{c^{k+1}}\, dc.$$

The expression for J in the general case where the x are not necessarily positive chance variables is

$$J = \Pi\{|\,x_r\,|\} \int_0^{\infty} \frac{F\{x_1/c, \ldots, x_k/c\}}{c^{k+1}}\, dc.$$

Small values of this are significant. The axial planes divide the k dimensional space into 2^k regions, and the probability of the sample point falling in any one of these regions is independent of the particular values of the parameters c. By considering these possibilities separately and using the result for the case of all positive chance variables, it is easy to see that the test is still unbiased. J is a homogeneous function of degree 0 in the x; hence, when H_1 is true, its distribution is independent of the particular common value of the scale parameters c. Therefore in determining this distribution we may take the common value to be 1.

From the corresponding result of § 1, it follows that an unbiased test of the hypothesis H_1', that

$$c_1 = c_2 = \ldots = c_k = 1,$$

will be obtained by using

$$J' = \Pi\{|\,x_r\,|\}\,.\, F\{x_1, \ldots, x_k\},$$

small values of which are significant. Again, from the corresponding result of § 1 it follows that if the x are independent positive chance variables with elementary probability functions $c_r^{-1} f_r(x_r/c_r)$ such that $x f_r(x)$ is a unimodal function of x, then,

when H_1' is not true, the probability that $J' \geqq \theta$ increases as any c moves towards the value 1.

Tests of other hypotheses will be discussed in a later paper. The remainder of this paper will be devoted to a discussion of the most important application of the results of this section.

The problem of deciding whether k samples, each known to have been drawn from a normal population, have been drawn from populations with the same standard deviation has been discussed by Neyman & Pearson (1931), who have proposed a criterion λ_{H_1} for this purpose. Applying the method of this paper, we obtain a test which is the same as that proposed by Bartlett (1937, p. 273). Bartlett's test is therefore unbiased. It can be shown that the Neyman-Pearson test is unbiased only when the number in each sample is the same, in which case it is exactly equivalent to Bartlett's test.

A continuous chance variable which takes values from 0 to ∞ with elementary probability function

$$\frac{1}{\Gamma(m)} e^{-x} x^{m-1}$$

will be called a $\Gamma(m)$ variable. If its elementary probability function has the more general form

$$\frac{1}{c\Gamma(m)} e^{-x/c} \left(\frac{x}{c}\right)^{m-1},$$

we shall call it an *unscaled* $\Gamma(m)$ variable with scale parameter c. If

$$y_1, y_2, \ldots, y_n$$

are n numbers with mean $\bar{y}$, we shall call the expression

$$S = \sum_1^n (y_r - \bar{y})^2$$

their *squariance*.† We know that if the y are a sample of n values from a normal population of standard deviation σ, S/σ^2 is distributed like χ^2 with $n-1$ degrees of freedom, or, what is more convenient here, $\frac{1}{2}S/\sigma^2$ has a $\Gamma\{\frac{1}{2}(n-1)\}$ distribution, so that S is an unscaled $\Gamma\{\frac{1}{2}(n-1)\}$ variable with scale parameter $2\sigma^2$. Moreover, in this case, if the mean of the population is unknown, the whole of the information about σ supplied by the sample is contained in the value of S. If the mean of the population has the known value a,

$$\sum_1^n (y_r - a)^2$$

is a sufficient statistic for the estimation of σ; it is an unscaled $\Gamma(\frac{1}{2}n)$ variable with scale parameter $2\sigma^2$. Hence questions about variances of normal populations can

† If this word is objected to, I hope that someone will coin a better, for it is very convenient to have a name for this expression. The usual "sum of squares" is not precise in meaning and is awkward in use. As I have pointed out previously (Pitman, 1937, p. 215), it is the squariance of a sample of n that appears naturally in the mathematical theory, and not the variance S/n, or the "estimated variance" $S/(n-1)$. The last, however, becomes prominent in the applications.

be reduced to questions about scale parameters of unscaled Γ variables. In particular, if we have k samples with squariances

$$S_1, S_2, \ldots, S_k,$$

and if each sample has been drawn from a normal population of unknown mean and variance, the question "Are the variances of the normal populations all equal?" is equivalent to "Are the scale parameters of the unscaled Γ variables

$$S_1, S_2, \ldots, S_k$$

all equal?" The problem of answering this latter question we now proceed to consider.

3. Application to gamma variables

Suppose that
$$x_1, x_2, \ldots, x_k$$

are k observed numbers, that x_r is a value of an unscaled $\Gamma(m_r)$ variable with scale parameter c_r, and that the k chance variables are independent. We wish to test the hypothesis H_1, that
$$c_1 = c_2 = \ldots = c_k.$$

Put
$$M = \Sigma m.$$

Except where explicitly stated otherwise, all summations Σ and all products Π are to be taken over the k values of the operand, which will be written without an index. Thus
$$\Sigma m = \sum_{r=1}^{k} m_r, \quad \Pi\{x^m\} = \prod_{r=1}^{k} \{x_r^{m_r}\}.$$

The elementary probability function of the distribution of the x is

$$\frac{e^{-\Sigma(x/c)}\Pi\{(x/c)^{m-1}\}}{\Pi\{c\Gamma(m)\}}.$$

Hence
$$J = \frac{\Pi\{x^m\}}{\Pi\{\Gamma(m)\}}\int_0^\infty e^{-(\Sigma x)/c}\frac{dc}{c^{M+1}}$$
$$= \frac{\Gamma(M)}{\Pi\{\Gamma(m)\}}\frac{\Pi\{x^m\}}{(\Sigma x)^M}.$$

It is sometimes more convenient to deal with

$$K = \frac{\Pi\{x^m\}}{(\Sigma x)^M},$$

which differs from J only by a constant factor. The maximum value of K is

$$\frac{\Pi\{m^m\}}{M^M},$$

and we shall denote the logarithm of the ratio of this maximum to the value of K by L, so that

$$L = M\log(\Sigma x/M) - \Sigma\{m\log(x/m)\}.$$

It is necessarily positive or zero, and large values are significant.

For testing the hypothesis H_1', that

$$c_1 = c_2 = \ldots = c_k = 1,$$

the appropriate criterion is
$$J' = \frac{\Pi\{x^m\}\, e^{-\Sigma x}}{\Pi\{\Gamma(m)\}}.$$

We shall write
$$K' = e^{-\Sigma x}\, \Pi\{x^m\},$$

$$L' = \log(K'_{\max}) - \log K' = \Sigma x - M - \Sigma\{m \log(x/m)\}.$$

4. The distribution of K when H_1 is true

As explained above, when H_1 is true we may take the common value of the c to be 1. The probability that $K \geqq \kappa$ is

$$p = \frac{1}{\Pi\{\Gamma(m)\}} \int e^{-\Sigma x}\, \Pi\{x^{m-1}\}\, dx_1 \ldots dx_k$$

over the region where
$$\frac{\Pi\{x^m\}}{(\Sigma x)^M} \geqq \kappa.$$

Make the change of variables

$$\Sigma x = u,$$
$$x_r = u y_r, \quad (r = 1, 2, \ldots, k-1)$$

and write for convenience
$$y_k = 1 - \sum_1^{k-1} y_r,$$

so that
$$\Sigma y = 1 \quad \text{and} \quad x_k = u y_k.$$

Then integrating with respect to u, which is not involved in the boundary condition, we obtain for the value of p

$$\frac{\Gamma(M)}{\Pi\{\Gamma(m)\}} \int \Pi\{y^{m-1}\}\, dy_1 \ldots dy_{k-1} \text{ over } \Pi\{y^m\} \geqq \kappa.$$

As the maximum value of K is $\Pi\{(m/M)^m\}$, we are concerned only with values of κ less than this. It should also be noted that in the applications to normal variables only integral and half-integral values of $m_1, m_2, \ldots$ occur.

When $\kappa = 0$, $p = 1$, and the region of integration is

$$0 \leqq y_r \leqq 1, \quad (r = 1, 2, \ldots, k-1)$$
$$0 \leqq \sum_1^{k-1} y_r \leqq 1.$$

Therefore
$$\int \Pi\{y^{m-1}\}\, dy_1 \ldots dy_{k-1}$$

over this region is equal to
$$\frac{\Pi\{\Gamma(m)\}}{\Gamma(M)},$$

which may be denoted by $B(m_1, m_2, \ldots, m_k)$. It is the extension of the complete beta function,

$$B(m_1, m_2) = \frac{\Gamma(m_1)\,\Gamma(m_2)}{\Gamma(m_1 + m_2)} = \int_0^1 x^{m_1 - 1}(1 - x)^{m_2 - 1}\, dx.$$

When $k = 2$, the expression for p is

$$\frac{1}{B(m_1, m_2)}\int_{x_1}^{x_2} x^{m_1-1}(1-x)^{m_2-1}\,dx,$$

where $$x_1^{m_1}(1-x_1)^{m_2} = x_2^{m_1}(1-x_2)^{m_2} = \kappa.$$

In the tabulations in § 5 for $k = 2$, the value of x_1 is given.

When $k = 3$,

$$p = \frac{1}{B(m_1, m_2, m_3)}\int y_1^{m_1-1} y_2^{m_2-1}(1-y_1-y_2)^{m_3-1}\,dy_1\,dy_2$$

over $$y_1^{m_1} y_2^{m_2}(1-y_1-y_2)^{m_3} \geqq \kappa.$$

By the substitution $$y_1 = uv \qquad (0 \leqq u \leqq 1),$$

$$y_2 = u(1-v) \quad (0 \leqq v \leqq 1),$$

this becomes

$$p = \frac{1}{B(m_1, m_2, m_3)}\int u^{m_1+m_2-1}(1-u)^{m_3-1}\,v^{m_1-1}(1-v)^{m_2-1}\,du\,dv \quad \text{......(3)}$$

over $$u^{m_1+m_2}(1-u)^{m_3}\,v^{m_1}(1-v)^{m_2} \geqq \kappa.$$

If $m_1 = m_2 = m_3 = m$, and m is integral or half-integral, this is expressible in terms of complete elliptic integrals; but even when $m = 1$ the reduction to standard form is tedious and I have not carried it out. However, Nair (1939) has shown how to obtain the exact distribution in the general case

$$m_1 = m_2 = \ldots = m_k.$$

It is of some interest to determine the exact distribution in other cases in order to check the accuracy of the approximate distribution discussed in § 5. A case which happens to be fairly easy to handle is

$$m_1 = m_2 = m, \quad m_3 = 2m.$$

From (3)

$$p = \frac{1}{B(m, m, 2m)}\int u^{2m-1}(1-u)^{2m-1}\,v^{m-1}(1-v)^{m-1}\,du\,dv$$

over $$u^2(1-u)^2\,v(1-v) \geqq \kappa^{1/m}.$$

Putting $$u = \tfrac{1}{2}(1+x) \quad (-1 \leqq x \leqq 1),$$

$$v = \tfrac{1}{2}(1+y) \quad (-1 \leqq y \leqq 1),$$

we obtain

$$p = \frac{1}{2^{6m-2}\,B(m, m, 2m)}\int (1-x^2)^{2m-1}\,(1-y^2)^{m-1}\,dx\,dy$$

over $$(1-x^2)^2\,(1-y^2) \geqq 64\kappa^{1/m},$$

which is fairly easily expressible in terms of complete elliptic integrals when m is small and integral or half-integral. When $m = 1$, so that

$$m_1 = m_2 = 1, \quad m_3 = 2,$$

$$p = \tfrac{3}{8}\int (1-x^2)\,dx\,dy \quad \text{over} \quad (1-x^2)^2(1-y^2) \geqq 64\kappa$$

$$= \tfrac{3}{4}\int \sqrt{\{(1-x^2)^2 - 64\kappa\}}\,dx \quad \text{over} \quad (1-x^2)^2 \geqq 64\kappa$$

$$= \tfrac{3}{2}\int_0^a \sqrt{\{(a^2-x^2)(b^2-x^2)\}}\,dx,$$

where $$a^2 = 1 - 8\sqrt{\kappa}, \quad b^2 = 1 + 8\sqrt{\kappa}.$$

This reduces to

$$p = \tfrac{1}{2}(a^2+b^2)\int_0^a \sqrt{\frac{b^2-x^2}{a^2-x^2}}\,dx - \tfrac{1}{2}b^2(b^2-a^2)\int_0^a \frac{dx}{\sqrt{\{(a^2-x^2)(b^2-x^2)\}}}$$

$$= b\int_0^{\frac{1}{2}\pi} \sqrt{(1-t^2\sin^2\phi)}\,d\phi - 8b\sqrt{\kappa}\int_0^{\frac{1}{2}\pi} \frac{d\phi}{\sqrt{(1-t^2\sin^2\phi)}},$$

where $t = a/b$. In the table in § 5 the value of $\theta = \arcsin t$ is given in order that the results may more readily be checked from tables of the complete elliptic integrals.

5. The semi-invariants of L and the approximate distribution of L when H_1 is true

We shall denote the sth semi-invariant of a chance variable Z by $\lambda_s(Z)$. If x is a $\Gamma(m)$ variable,

$$\lambda_s(\log x) = G_s(m),$$

where $G_s(m)$ is the sth derivative with respect to m of

$$G(m) = \log \Gamma(m).$$

From this we can easily show that the semi-invariants of

$$L = M\log(\Sigma x/M) - \Sigma\{m\log(x/m)\}$$

are $$\lambda_1(L) = M\{G_1(M) - \log M\} - \Sigma\{m[G_1(m) - \log m]\},$$

$s > 1$, $$\lambda_s(L) = (-)^s\{\Sigma m^s G_s(m) - M^s G_s(M)\}.$$

Using the asymptotic expansion of $G_s(m)$ and neglecting terms of order $1/m_r^2$, we obtain

$$\lambda_s(L^*) = \frac{k-1}{2}\,\Gamma(s),$$

where $$L^* = \frac{L}{1+\alpha},$$

$$\alpha = \frac{1}{6(k-1)}\{\Sigma(1/m) - 1/M\}.$$

Hence L^* is approximately a $\Gamma\{\frac{1}{2}(k-1)\}$ variable. $2L^*$ is distributed approximately like χ^2 with $k-1$ degrees of freedom, which is Bartlett's result (1937, p. 274), with a different notation.

The corresponding result used for testing H_1' is that, when H_1' is true,

$$2L^* = \frac{2L'}{1+\alpha}$$

is distributed approximately like χ^2 with k degrees of freedom, where

$$\alpha = \frac{1}{6k}\Sigma(1/m).$$

In the tables below, P is the true probability of obtaining, when H_1 is true, a value of $2L^*$ as great as or greater than that shown. P' is the approximate probability calculated from the approximate distribution of $2L^*$. Although in practice we are concerned only with the upper tail of the distribution of $2L^*$, it is of some interest to see how the approximate distribution as a whole compares with the true distribution.† It will be noted that the approximation improves as α decreases. The meanings of the symbols x_1 and θ are explained in § 4.

TABLE I

Comparing true and approximate probability integrals

$m_1=m_2=\frac{1}{2},\ \alpha=\frac{1}{2}$				$m_1=m_2=1,\ \alpha=\frac{1}{4}$			
x_1	$2L^*$	P	P'	x_1	$2L^*$	P	P'
0·42	0·01729	0·8977	0·8954	0·475	0·00401	0·95	0·9495
0·35	0·06287	0·8060	0·8020	0·45	0·01608	0·90	0·8991
0·27	0·15850	0·6957	0·6905	0·35	0·15090	0·70	0·6977
0·15	0·44890	0·5064	0·5029	0·25	0·46029	0·50	0·4975
0·05	1·10715	0·2871	0·2927	0·15	1·07735	0·30	0·2993
0·01	2·15262	0·1275	0·1423	0·05	2·65717	0·10	0·1031
0·001	3·68164	0·0403	0·0550	0·025	3·72464	0·05	0·0536
0·0001	5·21670	0·0127	0·0224	0·005	6·26726	0·01	0·0123
0·000019	6·32320	0·0056	0·0119	0·0025	7·37228	0·005	0·0066
$m_1=1,\ m_2=2,\ \alpha=\frac{7}{36}$				$m_1=1,\ m_2=\infty,\ \alpha=\frac{1}{6}$			
x_1	$2L^*$	P	P'	x_1	$2L^*$	P	P'
0·31	0·00631	0·9371	0·9367	0·91	0·00739	0·9318	0·9315
0·30	0·01303	0·9097	0·9091	0·87	0·01588	0·9002	0·8997
0·22	0·16997	0·6817	0·6801	0·70	0·09717	0·7561	0·7553
0·16	0·45502	0·5016	0·5000	0·40	0·54221	0·4625	0·4616
0·095	1·07829	0·2997	0·2991	0·20	1·38761	0·2385	0·2388
0·035	2·53527	0·1096	0·1113	0·10	2·40444	0·1195	0·1210
0·016	3·78064	0·0496	0·0518	0·04	3·87237	0·0472	0·0491
0·004	6·06129	0·0122	0·0138	0·01	6·19744	0·0115	0·0128
0·002	7·21519	0·0061	0·0072	0·004	7·75793	0·0045	0·0053

† See Bishop & Nair (1939), where the results of a very thorough investigation of the upper tail of the distribution are given. The computations of the present paper were completed before the publication of Bishop & Nair's paper.

TABLE I (*cont.*)

$m_1=m_2=5,\ \alpha=\frac{1}{20}$				$m_1=m_2=1,\ m_3=2,\ \alpha=\frac{3}{16}$			
x_1	$2L^*$	P	P'	θ	$2L^*$	P	P'
0·49	0·003809	0·95081	0·95079	7°	0·10005	0·9517	0·9512
0·48	0·015252	0·90177	0·90171	10°	0·20321	0·9043	0·9034
0·43	0·188518	0·66432	0·66415	19°	0·71677	0·7004	0·6988
0·39	0·472482	0·49205	0·49185	27°	1·40868	0·4952	0·4944
0·31	1·485259	0·22306	0·22295	36°	2·42738	0·2958	0·2971
0·25	2·739828	0·09785	0·09788	50°	4·53262	0·1001	0·1037
0·21	3·905482	0·04806	0·04813	57°	5·88759	0·0493	0·0527
0·14	6·958488	0·00828	0·00834	69°	9·02451	0·0093	0·0110
0·11	8·928707	0·00277	0·00281	75°	11·32588	0·0027	0·0035

6. Applications

Suppose that we have k samples, each drawn from a normal population of unknown mean and variance, and that we wish to test the hypothesis that the variances of the normal populations are all the same. Denote the number of members of the rth sample by n_r+1, and the squariance of this sample by S_r; then S_r is an unscaled $\Gamma(\frac{1}{2}n_r)$ variable, and the hypothesis to be tested is equivalent to the hypothesis H_1 applied to the variables S_r. Thus the appropriate criterion is†

$$L = \tfrac{1}{2}N\log(\Sigma S/\tfrac{1}{2}N) - \Sigma\{\tfrac{1}{2}n\log(S/\tfrac{1}{2}n)\},$$

where
$$N = \Sigma n.$$

A large value of L is significant.

$$2L = N\log(\Sigma S/N) - \Sigma\{n\log(S/n)\},$$

$$2L^* = \frac{2L}{1+\alpha},$$

where
$$\alpha = \frac{1}{6(k-1)}\left\{\Sigma\frac{1}{\frac{1}{2}n} - \frac{1}{\frac{1}{2}N}\right\} = \frac{\Sigma(1/n)-1/N}{3(k-1)}.$$

When H_1 is true, $2L^*$ is distributed approximately like χ^2 with $k-1$ degrees of freedom.‡ From the tabulations in § 5 it is evident that even with very small samples the approximation is sufficiently good for us to use it to determine

† The difference between this expression, L, and the function L_1 as defined by Neyman & Pearson should be noted. The relation is referred to on p. 212 below as well as being discussed by Bishop & Nair (1939).

‡ Cf. Bartlett (1937, p. 274).

whether an observed value of L is significant or not. It may be noted that if we put

$$v_r = S_r/n_r,$$

the "estimated variance" derived from sample r, the expression for $2L$ is

$$2L = N\log \bar{v} - \Sigma(n \log v),$$

where

$$\bar{v} = \Sigma nv/N.$$

Hence L has its minimum value 0 when

$$v_1 = v_2 = \ldots = v_k,$$

and $\partial L/\partial v_r$ has the same sign as $v_r - \bar{v}$, and therefore L increases as the v diverge more and more from a common value. Thus the L test may be regarded as answering the question "Are the estimated variances $v_1, v_2, \ldots, v_k$ significantly different?"

If we wish to test the hypothesis that the variance of each population is unity, we obtain the appropriate criterion from the fact that this hypothesis is equivalent to the hypothesis H' applied to the variables $\frac{1}{2}S_r$. The criterion is therefore

$$L' = \Sigma\tfrac{1}{2}S - \tfrac{1}{2}N - \Sigma\{\tfrac{1}{2}n\log(\tfrac{1}{2}S/\tfrac{1}{2}n)\},$$

and we have

$$2L' = \Sigma S - N - \Sigma\{n\log(S/n)\},$$

$$2L^* = \frac{2L'}{1+\alpha}, \quad \alpha = \frac{\Sigma(1/n)}{3k}.$$

When H' is true, $2L^*$ is distributed approximately like χ^2 with k degrees of freedom.

It should be noted that the field of application of the L test is quite different from that of Fisher's z test, which is used in the analysis of variance. When, in the analysis of variance, we test for significance of treatment differences or significance of an interaction, the question we put is not "Are these several estimated variances significantly different?" but "Is this particular estimate, or are any of this particular group, significantly greater than that particular estimated variance (residual variance)?" The truth is that, when engaged in the analysis of variance, we are interested not in variances of normal populations but in *means* of populations. A straightforward treatment of an analysis of variance problem by the method of likelihood leads to a criterion function which, when the null hypothesis is true (treatments without real differences of effect, or no interaction, as the case may be), is certainly the ratio of two estimated variances; but the question raised is not "Is this ratio significantly different from 1?" but "Is this ratio significantly *greater* than 1?"† We are led to Fisher's z test or an equivalent, and not to the L test. It is true that in the analysis of variance we have certain squariances S_1, S_2, ... which, when the null hypothesis is true, are unscaled gamma variables with equal scale parameters, and so the corresponding L will have the distribution discussed above; but this would not justify the use of the value of L as a test for the null hypothesis. In devising a satisfactory test we must consider the state of

† That being so, it may be remarked in passing, to speak of a "negative interaction" in a way which implies that it has some significance is a breach of statistical good manners, it is questioning the referee's decision after having agreed to accept it as final.

affairs not only when the hypothesis to be tested is true but also when it is false. In this case, when the null hypothesis is false, some at least of S_1, S_2, ... are not unscaled gamma variables but have a different distribution, and so the proof that the test is unbiased no longer holds. When used outside the field in which it naturally arises the L test loses its theoretical satisfactoriness. If, for example, the totals for the various treatments were identical, the L test applied to variance due to treatments and residual variance would judge this significant, i.e. as indicating that the treatments produce real differences in effect, while the z test would rightly judge the result as non-significant. Improbable as such a result is, it is more probable when the null hypothesis is true than when it is false. It would seem that the L test must lose power by judging as significant results in which the variance due to treatments is small compared with the residual variance, and that the z test must be more powerful in this field than the L test, i.e. more likely to reject the null hypothesis when it is false. Wishart (1938) proposes to use the L test as a supplementary test in experiments of factorial design when the z test has given a non-significant result for treatments as a whole, but it seems unlikely that this combination of the tests will be as powerful as the z test used alone.

7. The Neyman-Pearson test

The Neyman-Pearson criterion function for testing the hypothesis that the k normal samples have been drawn from populations with the same variance is (in our notation)

$$\lambda_{H_1} = \frac{\Pi\{[S/(n+1)]^{\frac{1}{2}(n+1)}\}}{\{\Sigma S/(N+k)\}^{\frac{1}{2}(N+k)}},\dagger$$

small values of this being significant. Remembering that S_r is an unscaled $\Gamma(\frac{1}{2}n_r)$ variable, we see that this is equivalent to using the criterion function

$$K' = \frac{\Pi\{x^{m+\frac{1}{2}}\}}{(\Sigma x)^{M+\frac{1}{2}k}}$$

instead of

$$K = \frac{\Pi\{x^m\}}{(\Sigma x)^M}$$

for testing the hypothesis H_1 applied to the variables $x_1, x_2, \ldots, x_k$ of § 2.

Denoting by p' the probability that $K' \geqq \kappa'$, we can easily show that when

$$c_1 = c_2 = \ldots = c_k = c,$$

$$\frac{\partial p'}{\partial c_1} = \frac{1}{c\Pi\{\Gamma(m)\}} \int (z_1 - m_1)\, e^{-\Sigma z}\, \Pi\{z^{m-1}\}\, dz_1 \ldots dz_k$$

over the region

$$\frac{\Pi\{z^{m+\frac{1}{2}}\}}{(\Sigma z)^{M+k/2}} \geqq \kappa' \qquad \ldots\ldots(4)$$

† Neyman & Pearson have denoted $\lambda_{H_1}^{2/(N+k)}$ by L_1.

From consideration of the derivative with respect to z_1 of the function on the left-hand side of (4) it is obvious that a line parallel to the z_1 axis, which cuts the boundary of the conical region (4), cuts it in two points and all points between these lie within the region (4). Let ζ, ζ' ($\zeta<\zeta'$) be the z_1 co-ordinates of these points—they will be functions of $z_2, z_3, \ldots, z_k$. Consider

$$\int_{\zeta}^{\zeta'} z_1 e^{-\Sigma z}\Pi\{z^{m-1}\}\,dz_1.$$

Integrating by parts, we obtain

$$\Big[-e^{-\Sigma z}z_1\,\Pi\{z^{m-1}\}\Big]_{\zeta}^{\zeta'}+m_1\int_{\zeta}^{\zeta'} e^{-\Sigma z}\Pi\{z^{m-1}\}\,dz_1$$
$$=\left[\frac{-e^{-\Sigma z}z_1(\Sigma z)^{M+k/2}\kappa'}{\Pi\{z^{\frac{1}{2}}\}}\right]_{\zeta}^{\zeta'}+m_1\int_{\zeta}^{\zeta'} e^{-\Sigma z}\Pi\{z^{m-1}\}\,dz_1,$$

since when $z_1=\zeta$ or ζ' the equality in (4) is satisfied. Hence

$$\int_{\zeta}^{\zeta'}(z_1-m_1)\,e^{-\Sigma z}\,\Pi\{z^{m-1}\}\,dz_1=\left[\frac{-e^{-\Sigma z}z_1(\Sigma z)^{M+k/2}\kappa'}{\Pi\{z^{\frac{1}{2}}\}}\right]_{\zeta}^{\zeta'}$$
$$=\kappa'\int_{\zeta}^{\zeta'}\frac{e^{-\Sigma z}z_1(\Sigma z)^{M+k/2}}{\Pi\{z^{\frac{1}{2}}\}}\left\{1-\frac{M+\frac{1}{2}k}{\Sigma z}+\frac{1}{2z_1}\right\}dz_1.$$

Thus, when $c_1=c_2=\ldots=c_k=c$, we have

$$\frac{\partial p'}{\partial c'}=\frac{\kappa'}{c\Pi\{\Gamma(m)\}}\int\frac{e^{-\Sigma z}z_1(\Sigma z)^{M+k/2}}{\Pi\{z^{\frac{1}{2}}\}}\left\{1-\frac{M+\frac{1}{2}k}{\Sigma z}+\frac{1}{2z_1}\right\}dz_1\ldots dz_k$$

over the region (4). By the method of § 4 we obtain

$$\frac{\partial p'}{\partial c_1}=\frac{\kappa'\Gamma(M)}{2c\Pi\{\Gamma(m)\}}\int_R\frac{1-ky_1}{\Pi\{y^{\frac{1}{2}}\}}\,dy_1\ldots dy_{k-1}$$

over the region R determined by

$$\Pi\{y^{m+\frac{1}{2}}\}\geqq\kappa',$$

where $\Sigma y=1$ as before. A necessary condition for the K' test to be unbiased is

$$\frac{\partial p'}{\partial c_r}=0,\quad(r=1,2,\ldots,k)$$

when $$c_1=c_2=\ldots=c_k,$$

and therefore for an unbiased test we must have

$$k\int_R\frac{y_r}{\Pi\{y^{\frac{1}{2}}\}}dy_1\ldots dy_{k-1}=\int_R\frac{1}{\Pi\{y^{\frac{1}{2}}\}}dy_1\ldots dy_{k-1}\quad(r=1,2,\ldots,k).\;\ldots\ldots(5)$$

It is obvious from symmetry and from the fact that $\Sigma y=1$ that (5) is true when

$$m_1=m_2=\ldots=m_k=m.$$

In fact the K' test is then equivalent to the K test since

$$K' = K^{(m+\frac{1}{2})/m},$$

and the K' test is therefore in that case unbiased. If the m are not all equal, suppose $m_1 > m_2$, then
$$a^{m_1+\frac{1}{2}}\, b^{m_2+\frac{1}{2}} \gtreqless b^{m_1+\frac{1}{2}}\, a^{m_2+\frac{1}{2}}$$

according as
$$a \gtreqless b.$$

Hence if P is any point in R with its y_1 co-ordinate less than its y_2 co-ordinate, the point P' obtained by interchanging the y_1 and y_2 co-ordinates of P will also lie in R, but if P is on the boundary of R (or sufficiently near it and inside R), and has its y_1 co-ordinate greater than its y_2 co-ordinate, the point P' will lie outside R.

Therefore
$$\int_R \frac{y_1 - y_2}{\Pi\{y^{\frac{1}{2}}\}}\, dy_1 \ldots dy_{k-1} > 0,$$

and so (5) cannot be true. Thus the K' test is biased when the m are not all equal. Considering the application to samples from normal populations, we see that the Neyman-Pearson test is biased unless the number in each sample is the same, in which case it is equivalent to the L test.

8. The distribution of K when H_1 is false

The probability that K is greater than or equal to κ is

$$p = \frac{1}{\Pi\{c^m \Gamma(m)\}} \int e^{-\Sigma(x/c)}\, \Pi\{x^{m-1}\}\, dx_1 \ldots dx_k$$

over the region where
$$\frac{\Pi\{x^m\}}{(\Sigma x)^M} \geqq \kappa,$$

which is equal to
$$\frac{1}{\Pi\{\Gamma(m)\}} \int e^{-\Sigma z}\, \Pi\{z^{m-1}\}\, dz_1 \ldots dz_k$$

over
$$\frac{\Pi\{c^m z^m\}}{(\Sigma cz)^M} \geqq \kappa.$$

By the method of §4 we obtain

$$p = \frac{\Gamma(M)}{\Pi\{\Gamma(m)\}} \int \Pi\{y^{m-1}\}\, dy_1 \ldots dy_{k-1}$$

over
$$\frac{\Pi\{c^m y^m\}}{(\Sigma cy)^M} \geqq \kappa,$$

where $\Sigma y = 1$. When $k = 2$, this can be evaluated by means of the tables of the incomplete beta function. It can be evaluated also in some simple cases of $k = 3$. An approximation to the distribution of L in the general case is being investigated.

SUMMARY

Tests of certain hypotheses concerning location and scale parameters are developed. These tests are unbiased and are applicable to chance variables with any continuous distributions. An application to the comparison of variances of normal variables yields Bartlett's test, which is thus shown to be unbiased. The Neyman-Pearson variances test is shown to be biased except when the samples are all of the same size.

REFERENCES

BARTLETT, M. S. (1937). *Proc. Roy. Soc.* A, **160**, 268–82.
BISHOP, D. J. & NAIR, U. S. (1939). *J. R. Statist. Soc. Suppl.* **6**, 89–99.
NAIR, U. S. (1939). *Biometrika*, **30**, 274–94.
NEYMAN, J. & PEARSON, E. S. (1931). *Bull. int. Acad. Cracovie*, A, 460–81.
—— —— (1936). *Statist. Res. Mem.* **1**, 1–37.
PITMAN, E. J. G. (1937). *Proc. Camb. Phil. Soc.* **33**, 212–22.
WISHART, J. (1938). *J. Agric. Sci.* **28**, 299–306.

NOTE

[As one of the authors of the original L_1 test and of the conception of "bias" in connexion with tests of statistical hypotheses, it is perhaps permissible for me to say that Prof. Neyman and I had for some time realized that the L_1 test was slightly biased when applied to samples of unequal size. The difference between this test and that put forward by M. S. Bartlett, and discussed more fully by Prof. Pitman above, may be expressed by saying that the Bartlett test weights the sums of squares with degrees of freedom where the L_1 test had used sample sizes as weights. Intuitionally the former weighting seems the more appropriate to take but, as in certain problems of estimation with which the reader will probably be familiar, the application of the principle of maximum likelihood, a form of which Neyman and I had used, introduces sample sizes rather than degrees of freedom. In 1936 B. L. Welch (*Statistical Research Memoirs*, **1**, 53) had suggested the use of a test function in the form of the ratio of a weighted geometric mean to a weighted arithmetic mean of sums of squares, the weights being adjustable at will; but it was Bartlett who definitely advocated the weighting with degrees of freedom.

In a Ph.D. Thesis for the University of London, completed in 1937, only a part of which has been published (U. S. Nair, 1939), Dr Nair showed that in a particular case with $k=3$ and sample sizes of 5, 10 and 15 the L_1 test was certainly biased. During the past winter he also sent me a mathematical proof showing that in the simplest case, with $k=2$, the Bartlett test was unbiased. Of two or three people investigating the problem, Prof. Pitman has however been the first to reach a complete solution. His proof is of great generality; not only does it cover the case of any number of samples or groups of data, but it is derived from a very interesting general approach not limited to the particular problem of testing for homogeneity of standard deviations among normally distributed variables. E. S. P.]

MISCELLANEA

(i) On the Calculation of the Cumulants of the χ-distribution

By N. L. JOHNSON and B. L. WELCH

In a recent investigation it has been found necessary to calculate the cumulants of the χ-distribution with considerable accuracy. This may be done, of course, by calculating the moments of χ about zero and then correcting in the usual way. It was found, however, that a substantial saving of labour may be effected by another method which will be indicated below. This method is essentially the same as that employed by K. Pearson (1915) to calculate the constants of the standard deviation distribution, being extended to give the fifth and sixth cumulants.

The distribution of χ with f degrees of freedom is given by

$$p(\chi)\,d\chi = \frac{1}{2^{\frac{1}{2}f-1}\Gamma(\frac{1}{2}f)}e^{-\frac{1}{2}\chi^2}\chi^{f-1}\,d\chi.$$

The kth moment of χ about zero is

$$\mu'_k = 2^{\frac{1}{2}k}\frac{\Gamma\{\frac{1}{2}(f+k)\}}{\Gamma(\frac{1}{2}f)},$$

whence it is seen that the even moments are integers and the odd moments integral multiples of μ'_1. The kth moment μ_k about the mean may therefore be expressed as a polynomial of the kth degree in μ'_1 with integral coefficients. Hence we obtain the following formulae for the first six cumulants:

$$\kappa_1 = \mu'_1,$$

$$\kappa_2 = f-\kappa_1^2,$$

$$\kappa_3 = (-2f+1)\kappa_1+2\kappa_1^3,$$

$$\kappa_4 = (-2f^2+2f)+(8f-4)\kappa_1^2-6\kappa_1^4,$$

$$\kappa_5 = (16f^2-16f+3)\kappa_1+(20-40f)\kappa_1^3+24\kappa_1^5,$$

$$\kappa_6 = (16f^3-24f^2+8f)+(-136f^2+136f-28)\kappa_1^2+(240f-120)\kappa_1^4-120\kappa_1^6.$$

By substituting $(f-\kappa_2)$ for κ_1^2 we may now obtain formulae for κ_3, κ_4, κ_5 and κ_6 involving only the first power of κ_1 and powers of κ_2 up to the third. Thus

$$\kappa_3 = \kappa_1(1-2\kappa_2),$$

$$\kappa_4 = -2f(1-2\kappa_2)+4\kappa_2-6\kappa_2^2,$$

$$\kappa_5 = \kappa_1[4f(1-2\kappa_2)+3-20\kappa_2+24\kappa_2^2],$$

$$\kappa_6 = -8f^2(1-2\kappa_2)-f(20-104\kappa_2+120\kappa_2^2)+(28\kappa_2-120\kappa_2^2+120\kappa_2^3).$$

The facts that the multiplier of the highest power of f is always a multiple of $(1-2\kappa_2)$ and that κ_2 tends to $\frac{1}{2}$ as f tends to infinity, suggest that some simplification may result by expressing

these cumulants in terms of powers of $(1-2\kappa_2)$ instead of powers of κ_2. Putting $(1-2\kappa_2)=\alpha$ we find

$$\kappa_3 = \kappa_1\alpha,$$

$$\kappa_4 = \tfrac{1}{2}-(2f-1)\alpha-\tfrac{3}{2}\alpha^2,$$

$$\begin{aligned}\kappa_5 &= \kappa_1[-1+(4f-2)\alpha+6\alpha^2]\\ &= \kappa_1[-2\kappa_4+3\alpha^2],\end{aligned}$$

$$\begin{aligned}\kappa_6 &= (2f-1)-(8f^2-8f-1)\alpha-15(2f-1)\alpha^2-15\alpha^3\\ &= 2(2f-1)\kappa_4+3\alpha-12(2f-1)\alpha^2-15\alpha^3.\end{aligned}$$

The calculation of the higher cumulants by means of these formulae is rendered very easy once κ_1 is known. For f small κ_1 may be calculated from the following formulae:

$$f \text{ even} \quad \kappa_1 = \sqrt{\frac{\pi}{2}}\,\frac{(f-1)(f-3)\ldots 3\cdot 1}{(f-2)(f-4)\ldots 2},$$

$$f \text{ odd} \quad \kappa_1 = \sqrt{\frac{2}{\pi}}\,\frac{(f-1)(f-3)\ldots 4\cdot 2}{(f-2)(f-4)\ldots 3\cdot 1}.$$

(To 15 decimal places $\sqrt{\frac{2}{\pi}} = 0\cdot79788\,45608\,02865,$

and $\sqrt{\frac{\pi}{2}} = 1\cdot25331\,41373\,15500.$)

For f large Stirling's expansion for the factorial may be used, giving

$$\log\kappa_1 = \tfrac{1}{2}\log f-\frac{1}{4f}+\frac{1}{24f^3}-\frac{1}{20f^5}+\frac{17}{112f^7}+\ldots,$$

$$\kappa_1 = \sqrt{f}\left\{1-\frac{1}{4f}+\frac{1}{32f^2}+\frac{5}{128f^3}-\frac{21}{2048f^4}-\frac{399}{8192f^5}+\frac{869}{65536f^6}+\ldots\right\}.$$

The above formulae allow the calculation of the cumulants very quickly with as much accuracy as is ever likely to be required, but the following short tables, giving only a few figures, will be useful for many practical purposes. Note that in the second table the cumulants are multiplied by powers of f in order to increase the accuracy of harmonic interpolation.

f	κ_1	κ_2	κ_3	κ_4	κ_5	κ_6
1	0·797885	0·363380	0·218014	0·114771	−0·004438	−0·152659
2	1·253314	0·429204	0·177460	0·045149	−0·037791	−0·068652
3	1·595769	0·453521	0·148340	0·022247	−0·029635	−0·029175
4	1·879971	0·465708	0·128935	0·012860	−0·021825	−0·014156
5	2·127692	0·472926	0·115210	0·008271	−0·016482	−0·007712
6	2·349964	0·477669	0·104953	0·005730	−0·012867	−0·004596
7	2·553231	0·481014	0·096954	0·004189	−0·010346	−0·002935
8	2·741625	0·483494	0·090506	0·003190	−0·008526	−0·001978

f	$24/f$	$\kappa_1/\sqrt{f}$	κ_2	$\kappa_3\sqrt{f}$	$\kappa_4 f^2$	$\kappa_5(\sqrt{f})^3$	$\kappa_6 f^3$
6	4	0·959369	0·477669	0·257082	0·20627	−0·18910	−0·9927
8	3	0·969311	0·483494	0·255989	0·20413	−0·19291	−1·0128
12	2	0·979406	0·489176	0·254427	0·20025	−0·19396	−1·0114
24	1	0·989640	0·494686	0·252418	0·19464	−0·19211	−0·9859
∞	0	1·000000	0·500000	0·250000	0·18750	−0·18750	−0·9375

REFERENCE

Pearson, Karl (1915). Appendix to Papers by "Student" and R. A. Fisher. Editorial. *Biometrika*, **10**, 522–9.

(ii) Note on Discriminant Functions

By B. L. WELCH, Ph.D.

Recent years have seen a widespread development of the methods of multivariate analysis. Of the problems which have been discussed one of the simplest is that of apportioning individuals with several measured characters into one or other of two completely specified population groups. Thus if $x_1, x_2, \ldots, x_q$ are the q characters measured upon each individual, and Π_1 and Π_2 denote the two possible populations, we suppose known the probability distributions $p_1 = p(x_1, x_2, \ldots, x_q \mid \Pi_1)$ and $p_2 = p(x_1, x_2, \ldots, x_q \mid \Pi_2)$. In particular it is frequently assumed that Π_1 and Π_2 have the same set of variances and covariances and differ only in the mean values of the characters. In this case R. A. Fisher (1936) has considered the problem of choosing the best linear function X of $x_1, x_2, \ldots, x_q$ to form the basis of classification of the individuals. The solution was obtained by maximizing the absolute value of the ratio of $E(X \mid \Pi_1) - E(X \mid \Pi_2)$ to the standard deviation of X. This is certainly the best discriminant function of any kind, whether linear or not, provided p_1 and p_2 are of the multivariate normal form.

However, without making any assumptions of normality or equality of variances and covariances, the problem of obtaining the best function to discriminate between two completely specified populations may still be solved. The function is simply the ratio of the two probability laws p_1/p_2. This is almost self-evident, but the following demonstrations may be useful.

Suppose in the first instance that it is possible to assess *a priori* probabilities ω_1 and ω_2 that an individual will belong to Π_1 or Π_2 respectively. This was possible, for instance, in the problem of sexing human mandibles, discussed by E. S. Martin (1936). Then, if measurements $x_1, x_2, \ldots, x_q$ are made on the individual, the *a posteriori* probabilities that it will belong to Π_1 or Π_2 will be $\omega_1 p_1/(\omega_1 p_1 + \omega_2 p_2)$ and $\omega_2 p_2/(\omega_1 p_1 + \omega_2 p_2)$, respectively. Now if it is equally important that an individual really belonging to Π_1 should not be classified as belonging to Π_2 and vice versa, then we should assign it to Π_1 provided the *a posteriori* probability $\omega_1 p_1/(\omega_1 p_1 + \omega_2 p_2)$ is greater than $\frac{1}{2}$. Any other assignation would increase the overall chance of misclassifications. We should therefore classify into Π_1 when

$$\frac{p_1}{p_2} > \frac{\omega_2}{\omega_1}. \qquad \ldots\ldots(1)$$

Whatever the prior probabilities the discriminant function to be calculated from the

observed measurements is therefore the ratio of the probability laws; the criterion level to which this function is to be referred does, however, depend on the prior probabilities.

If it is not possible or even appropriate to assess prior probabilities then the determination of a rule of classification must be made to depend on other conditions. For instance, it may be required (*a*) that if an individual really belongs to Π_1 there should be the same chance of it being misclassified as there would be if it belonged to Π_2 and (*b*) that this common chance should be a minimum. If $x_1, x_2, \ldots, x_q$ be represented by a point in q-dimensional space, any rule of classification involves the choice of a region R in this space such that an individual falling into R will be classified into Π_1. The above conditions (*a*) and (*b*) may then be written

$$\int_R p_2\, d\tau = 1 - \int_R p_1\, d\tau, \qquad \ldots\ldots(2)$$

and

$$\int_R p_2\, d\tau \text{ to be a minimum.} \qquad \ldots\ldots(3)$$

The problem is therefore to minimize $\int_R p_2\, d\tau$ subject to the condition that $\int_R (p_2+p_1)\, d\tau$ is equal to unity. A straightforward application of the calculus of variations shows that R consists of points such that

$$p_1/p_2 > k, \qquad \ldots\ldots(4)$$

where k is chosen to satisfy (2). The discriminant function is the same as before although the criterion level will in general be different.

Whatever the conditions which are imposed, provided they are formulated on a probability basis, it would appear that we shall be led to the same discriminant function. To take a further instance we may mention that in discussing the theory of testing statistical hypothesis, J. Neyman & E. S. Pearson (1932) have considered the problem of testing the hypothesis that a sample comes from a certain population when there is only one possible alternative population. They imposed the conditions (*a*) that the chance of rejection when the hypothesis is true should be a small specified probability ϵ, and (*b*) that the chance of rejection when the alternative hypothesis is true should be a maximum. The solution depended only on the ratio of the probability laws of the sample on the two hypotheses. This ratio has been termed by Neyman & Pearson a "likelihood ratio" and sometimes a "test criterion". It seems to me that "discriminant function" is a better term for this ratio and that "criterion" is best reserved for the particular value of the function which is considered critical. This value will, as we have seen, depend on the manner in which the problem of deciding between the two populations is formulated.

When the populations Π_1 and Π_2 are normal multivariate, differing only in their mean values, it is easy to see that the ratio p_1/p_2 leads to the same linear function of $x_1, x_2, \ldots, x_q$ as is obtained by maximizing $\{E(X \mid \Pi_1) - E(X \mid \Pi_2)\}^2/\sigma_X^2$. For if

$$p(x_1, x_2, \ldots, x_q \mid \Pi_t) = \left(\frac{1}{\sqrt{(2\pi)}}\right)^q \frac{1}{W^{\frac{1}{2}}} \exp\left[-\tfrac{1}{2} \sum_{i,j} W^{ij}(x_i - \theta_{it})(x_j - \theta_{jt})\right] \quad (t = 1, 2), \qquad \ldots\ldots(5)$$

where W is the determinant of the covariances and W^{ij} are the elements of the reciprocal matrix, then

$$\log \frac{p_1}{p_2} = -\tfrac{1}{2} \sum_{i,j} W^{ij}\{(x_i - \theta_{i1})(x_j - \theta_{j1}) - (x_i - \theta_{i2})(x_j - \theta_{j2})\}$$

$$= \sum_{i,j} W^{ij}(\theta_{j1} - \theta_{j2}) x_i - \tfrac{1}{2} \sum_{i,j} W^{ij}(\theta_{i1}\theta_{j1} - \theta_{i2}\theta_{j2}).$$

$\sum_{i,j} W^{ij}(\theta_{j1} - \theta_{j2}) x_i$ is the same linear function as is obtained from the other approach (R. A. Fisher, 1936).

REFERENCES

Fisher, R. A. (1936). "The use of multiple measurements in taxonomic problems." *Ann. Eugen., Lond.*, 7, 179–88.

Martin, E. S. (1936). "A study of an Egyptian series of mandibles, with special reference to mathematical methods of sexing." *Biometrika*, **28**, 149–72.

Neyman, J. & Pearson, E. S. (1932). "On the problem of the most efficient tests of statistical hypotheses." *Philos. Trans.* A, **231**, 289–337.

(iii) Principles of Genetics. By E. W. Sinnott and L. C. Dunn. Fourth Edition. London: McGraw-Hill Publishing Company, 1939. Price 21*s*.

This book, which has gone through three editions since 1925, needs no recommendation to geneticists. It is so good on the whole that a short review can be mainly concerned with its defects. Mendel did not use red-flowered peas as stated on p. 41, but purple-flowered (*rot-violett*) and, if he had crossed red and white, would probably have got a purple F_1. The treatment of biometrical methods on pp. 138–51 is unsatisfactory. For example, the account of the significance of a difference of means takes no account of "Student's" work, and is therefore just 31 years out-of-date. There is no need to point out the serious nature of the errors which may arise from this neglect. And since Bateson's original definition of genetics included a study of populations, I cannot but feel that the authors have devoted too little space to this topic. Nevertheless, as an elementary account of modern genetics, the book can be strongly recommended.

J. B. S. H.

(iv) Corrections to formulae in papers on the moments of χ^2. By J. B. S. Haldane.

The following errata should be noted:

Biometrika, **29**, 138, equations (3):

For $\mu_3 = \kappa_3 = 8 + 2(11k - 6)\, s^{-1} + (k^2 - 30k + 120)\, s^{-2}$,

read $\mu_3 = \kappa_3 = 8 + 2(11k - 56)\, s^{-1} + (k^2 - 30k + 120)\, s^{-2}$.

Biometrika, **29**, 390, antepenultimate line:

For $\kappa_6 = 3840 + 21{,}300m^{-1} + 249{,}600m^{-2} + 69{,}160m^{-3} + 2004m^{-4} + m^{-5}$,

read $\kappa_6 = 3840 + 124{,}800m^{-1} + 249{,}600m^{-2} + 69{,}160m^{-3} + 2004m^{-4} + m^{-5}$.

Biometrika, **29**, 391, equations (1):

For $\kappa_6 = 3840n + 21{,}300R_1 + 249{,}600R_2 + 69{,}160R_3 + 2004R_4 + R_5$,

read $\kappa_6 = 3840n + 124{,}800R_1 + 249{,}600R_2 + 69{,}160R_3 + 2004R_4 + R_5$.

And line 20:

For $+5814{\cdot}04n^2 + 18{,}640{\cdot}492n$,

read $+5814{\cdot}04n^2 + 39{,}340{\cdot}492n$.

J. B. S. H.

ON GENERALIZED ANALYSIS OF VARIANCE. (I)

BY P. L. HSU

1. *The Wilks-Lawley hypothesis.* Suppose that, as a result of some random sampling, we are in possession of $p(n_1+n)$ quantities, where $n \geqslant p$, calculated from the observational data. Calling these y_{ir} and $z_{ir'}$ ($i=1,2,\ldots,p$; $r=1,2,\ldots,n_1$; $r'=1,2,\ldots,n$), we assume that in repeated sampling they follow the probability distribution

$$\text{Const.}\times\exp\left\{-\frac{1}{2}\sum_{i,j=1}^{p}\alpha_{ij}\sum_{r=1}^{n_1}(y_{ir}-\eta_{ir})(y_{jr}-\eta_{jr})-\frac{1}{2}\sum_{i,j=1}^{p}\alpha_{ij}\sum_{r=1}^{n}z_{ir}z_{jr}\right\}\Pi\,dy\,dz, \quad (1)$$

and that we have no previous knowledge of the values of the η_{ir}. Our problem is to test the hypothesis H_0:

$$\eta_{ir}=0 \quad \text{for} \quad i=1,2,\ldots,p;\ r=1,2,\ldots,n_1. \qquad (H_0)$$

For $p=1$ the hypothesis is that to which every linear hypothesis* may be reduced, and the test amounts to that ordinarily employed in the analysis of variance. For $n_1=1$ the problem calls for Hotelling's generalized "Student's" test.†

It should be emphasized that the y_{ir}, z_{ir} and η_{ir} will not usually correspond to the original physical observations and physical constants belonging to the sampled populations, but are so derived from them that (1) is true and that H_0 is equivalent to some hypothesis regarding population constants that we want to test. In another paper we shall deal with a general class of "linear hypothesis" on multivariate normal means, showing how all of them can be reduced by a rotation of the sample space to the "canonical" H_0. Here we content ourselves with the following example.

Example. Case of k samples. $x_{i\nu t}$ ($i=1,2,\ldots,p$; $\nu=1,2,\ldots,k$; $t=1,2,\ldots,m_\nu$) are k samples drawn respectively from k p-variate normal populations all of which have the same set of variances and covariances. Let the population means be $\xi_{i\nu}$ ($i=1,2,\ldots,p$; $\nu=1,2,\ldots,k$). Let

$$\sum_{\nu=1}^{k}m_\nu=M, \quad \bar{\xi}_i=\frac{1}{M}\sum_{\nu=1}^{k}m_\nu\xi_{i\nu},$$

$$\bar{x}_{i\nu}=\frac{1}{m_\nu}\sum_{t=1}^{m_\nu}x_{i\nu t}, \quad \bar{x}_i=\frac{1}{M}\sum_{\nu=1}^{k}m_\nu\bar{x}_{i\nu},$$

$$s_{ij\nu}=\sum_{t=1}^{m_\nu}(x_{i\nu t}-\bar{x}_{i\nu})(x_{j\nu t}-\bar{x}_{j\nu}),$$

$$(i,j=1,2,\ldots,p;\ \nu=1,2,\ldots,k).$$

* Kolodziezczyk (1935), Tang (1938).

† Hotelling (1931), Hsu (1938), Bose and Roy (1938)

Suppose that we want to test the hypothesis:

$$\xi_{i1} = \xi_{i2} = \ldots = \xi_{ik} \quad \text{for} \quad i = 1, 2, \ldots, p.$$

Then it is known that if we call y_{ir} and z_{ir} certain linear functions of the $x_{i\nu t}$ and η_{ir} certain linear functions of the $\xi_{i\nu}$, (1) will hold true and the hypothesis under test is equivalent to H_0. Thus reducing the problem, we have

$$n_1 = k - 1, \quad n = M - k,$$

$$\sum_{r=1}^{n_1} y_{ir} y_{jr} = \sum_{\nu=1}^{k} m_\nu (\bar{x}_{i\nu} - \bar{x}_i)(\bar{x}_{j\nu} - \bar{x}_j), \quad \sum_{r=1}^{n} z_{ir} z_{jr} = \sum_{\nu=1}^{k} s_{ij\nu},$$

$$\sum_{r=1}^{n_1} \eta_{ir} \eta_{jr} = \sum_{\nu=1}^{k} m_\nu (\xi_{i\nu} - \bar{\xi}_i)(\xi_{j\nu} - \bar{\xi}_j),$$

$$(i, j = 1, 2, \ldots, p).$$

In particular, if, $k = 2$ then $n_1 = 1$, and we may drop the second index of y and η. We have then

$$y_i y_j = \frac{m_1 m_2}{m_1 + m_2} (\bar{x}_{i1} - \bar{x}_{i2})(\bar{x}_{j1} - \bar{x}_{j2}),$$

$$\sum_{r=1}^{n} z_{ir} z_{jr} = \sum_{t=1}^{m_1} (x_{i1t} - \bar{x}_{i1})(x_{j1t} - \bar{x}_{j1}) + \sum_{t=1}^{m_2} (x_{i2t} - \bar{x}_{i2})(x_{j2t} - \bar{x}_{j2}),$$

$$\eta_i \eta_j = \frac{m_1 m_2}{m_1 + m_2} (\xi_{i1} - \xi_{i2})(\xi_{j1} - \xi_{j2}),$$

$$(i, j = 1, 2, \ldots, p).$$

The hypothesis in the above example has been studied by Wilks (1932), while an attempt to test the general hypothesis H_0 was made by Lawley (1938). Both assumed no prior knowledge of the values of the α_{ij}. We propose to call H_0 the Wilks-Lawley hypothesis.

2. *Case where the α_{ij} are known, generalizations of Mahalanobis's distance.* From now on we shall write

$$a_{ij} = \sum_{r=1}^{n_1} y_{ir} y_{jr}, \quad b_{ij} = \sum_{r=1}^{n} z_{ir} z_{jr},$$

$$\psi_{ij} = \sum_{r=1}^{n_1} \eta_{ir} \eta_{jr}, \quad \Psi = \sum_{i,j=1}^{p} \alpha_{ij} \psi_{ij},$$

$$(i, j = 1, 2, \ldots, p).$$

In order to test H_0 when the α_{ij} are known, we calculate the likelihood ratio* and get

$$-2 \log(\text{likelihood ratio}) = \sum_{i,j=1}^{p} \alpha_{ij} a_{ij} = S, \text{ say}$$

and decide to reject H_0 if the value of S exceeds some fixed constant, chosen so as to fix at some desired level the risk of rejecting H_0 when it is true.

* For the definition and interpretation of the term see Neyman & Pearson (1928).

Theorem 1. *For arbitrary values of the η_{ir} the distribution of S is that of the sum of pn_1 non-central squares:**

$$2^{-\frac{1}{2}pn_1} S^{\frac{1}{2}pn_1-1} e^{-\frac{1}{2}(S+\Psi)} \left\{ \sum_{h=0}^{\infty} \frac{\Psi^h S^h}{4^h h! \Gamma(h+\frac{1}{2}pn_1)} \right\} dS. \tag{2}$$

In particular, if H_0 is true, then $\Psi = 0$ and S follows the χ^2 distribution with pn_1 degrees of freedom.

Proof. Let the sets of variables $(y_{1r}, y_{2r}, \ldots, y_{pr})$, for $r = 1, 2, \ldots, n_1$, be subject to the same linear transformation such that, calling the new variables $(u_{1r}, u_{2r}, \ldots, u_{pr})$,

$$\sum_{i,j=1}^{p} \alpha_{ij} y_{ir} y_{jr} = \sum_{i=1}^{p} u_{ir}^2, \quad (r = 1, 2, \ldots, n_1).$$

This is possible because the matrix $\|\alpha_{ij}\|$ is positive definite. Let μ_{ir} be the same linear function of the η's as u_{ir} is of the y's. Then

$$S = \sum_{i=1}^{p} \sum_{r=1}^{n_1} u_{ir}^2 \tag{3}$$

and the u_{ir} follow the distribution

$$(2\pi)^{-\frac{1}{2}pn_1} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{p} \sum_{r=1}^{n_1} (u_{ir} - \mu_{ir})^2 \right\} \Pi du. \tag{4}$$

From (3) and (4) follows the result (2),† as we have

$$\sum_{i=1}^{p} \sum_{r=1}^{n_1} \mu_{ir}^2 = \sum_{i,j=1}^{p} \alpha_{ij} \sum_{r=1}^{n_1} \eta_{ir} \eta_{jr} = \Psi.$$

From (2) we get
$$\mathscr{E}(S) = \Psi + pn_1, \tag{5}$$

$$\mathscr{E}(S^2) = \Psi^2 + 2(pn_1+2)\Psi + pn_1(pn_1+2),$$

whence
$$\sigma^2(S) = 4\Psi + 2pn_1.$$

If these general results are applied to the above example of k samples, we have

$$S = \sum_{i,j=1}^{p} \alpha_{ij} \sum_{\nu=1}^{k} m_\nu (\bar{x}_{i\nu} - \bar{x}_i)(\bar{x}_{j\nu} - \bar{x}_j) = S_k,$$

$$\Psi = \sum_{i,j=1}^{p} \alpha_{ij} \sum_{\nu=1}^{k} m_\nu (\xi_{i\nu} - \bar{\xi}_i)(\xi_{j\nu} - \bar{\xi}_j) = \Psi_k,$$

say, and the distribution

$$2^{-\frac{1}{2}p(k-1)} S_k^{\frac{1}{2}p(k-1)-1} e^{-\frac{1}{2}(S_k+\Psi_k)} \left\{ \sum_{h=0}^{\infty} \frac{\Psi_k^h S_k^h}{4^h h! \Gamma(h+\frac{1}{2}p(k-1))} \right\} dS_k. \tag{6}$$

For $k = 2$ we have

$$S_2 = \frac{m_1 m_2}{m_1+m_2} \sum_{i,j=1}^{p} \alpha_{ij} (\bar{x}_{i1} - \bar{x}_{i2})(\bar{x}_{j1} - \bar{x}_{j2}),$$

$$\Psi_2 = \frac{m_1 m_2}{m_1+m_2} \sum_{i,j=1}^{p} \alpha_{ij} (\xi_{i1} - \xi_{i2})(\xi_{j1} - \xi_{j2}),$$

* Fisher (1928), p. 669. † Fisher (1928), Tang (1938), pp. 138–9.

and the distribution

$$2^{-\frac{1}{2}p} S_2^{\frac{1}{2}p-1} e^{-\frac{1}{2}(S_2+\Psi_2)} \left\{ \sum_{h=0}^{\infty} \frac{\Psi_2^h S_2^h}{4^h h! \Gamma(h+\frac{1}{2}p)} \right\} dS_2. \tag{7}$$

The quantity Ψ_2 is equal to

$$\frac{pm_1 m_2}{m_1+m_2} \Delta,$$

where Δ is Mahalanobis's "distance"* between two populations having the same variances and covariances weighted according to m_1 and m_2. The statistic

$$\frac{m_1+m_2}{pm_1 m_2} S_2 - \frac{1}{m_1} - \frac{1}{m_2},$$

which, according to (5), is an unbiassed estimate of Δ, is called the D^2 statistic and the distribution (7) has already been obtained.† These concepts and formulae are completely generalized here by S_k, Ψ_k and the formulae (5) and (6) for the case $k > 2$.

3. *Tests suggested by Wilks and Lawley: Generalized* E^2, $1-E^2$ *and* $E^2(1-E^2)^{-1}$. From now on we shall assume complete ignorance of the values of the α_{ij}. In the case $p = 1$ the following functions are well known:

$$E^2 = \sum_{r=1}^{n_1} y_r^2 \Big/ \left(\sum_{r=1}^{n_1} y_r^2 + \sum_{r=1}^{n} z_r^2 \right),$$

$$1-E^2 = \sum_{r=1}^{n} z_r^2 \Big/ \left(\sum_{r=1}^{n_1} y_r^2 + \sum_{r=1}^{n} z_r^2 \right),$$

$$E^2/(1-E^2) = \sum_{r=1}^{n_1} y_r^2 \Big/ \sum_{r=1}^{n} z_r^2.$$

Wilks, while studying the k-sample case,‡ suggested the functions U and W respectively as generalizations of E^2 and $1-E^2$. In our general case these are defined as

$$U = \frac{|a_{ij}|}{|a_{ij}+b_{ij}|} \quad \text{if } n_1 \geqslant p, \tag{8}$$

and

$$W = \frac{|b_{ij}|}{|a_{ij}+b_{ij}|}. \tag{9}$$

Both are ratios of two determinants. The above definition for U based on the idea of a generalized E^2 breaks down if $n_1 < p$, as then $|a_{ij}|$ vanishes identically. This difficulty can be met if we define U as the product of the non-identically vanishing roots of the determinantal equation $|a_{ij} - \theta(a_{ij}+b_{ij})| = 0$. If $n_1 \geqslant p$, this definition evidently coincides with (8).

* Mahalanobis (1936). † Bose (1936). ‡ Wilks (1932).

As a generalization of $E^2(1-E^2)^{-1}$ we take, instead of UW^{-1}, the function V suggested by Lawley* and defined as

$$V = \sum_{i,j=1}^{p} b^{ij} a_{ij}, \tag{10}$$

where b^{ij} is the element (i, j) of the reciprocal matrix $\|b_{ij}\|^{-1}$.

If we denote by $\theta_1, \theta_2, \ldots, \theta_{l_1}$, where l_1 is the smaller of the integers p and n_1, the non-identically vanishing roots of the equation

$$|a_{ij} - \theta(a_{ij}+b_{ij})| = 0,$$

then, according to (9) and (10) and the new definition of V, we have

$$U = \prod_{i=1}^{l_1} \theta_i, \tag{11}$$

$$W = \prod_{i=1}^{l_1} (1-\theta_i), \tag{12}$$

$$V = \sum_{i=1}^{l_1} \frac{\theta_i}{1-\theta_i}. \tag{13}$$

W and V are test functions for H_0 suggested respectively by Wilks and Lawley. If we use W (or V), we reject H_0 if W is smaller (or V is greater) than some prescribed constant. In the following sections we shall study the distributions of U, V and W in repeated sampling.

4. *Case when H_0 is true.* If H_0 is true, then, putting all the $\eta_{ir} = 0$ in (1), we get the parent distribution

$$\text{const.} \times \exp\left\{-\frac{1}{2}\sum_{i,j=1}^{p} \alpha_{ij}(a_{ij}+b_{ij})\right\} \Pi\, dy\,dz. \tag{14}$$

The following theorem has been proved elsewhere.†

THEOREM 2. *Let l_1 be the smaller, and l_2 the larger, of the integers p and n_1. Let $\theta_1, \theta_2, \ldots, \theta_{l_1}$ be the non-identically vanishing roots of the determinantal equation*

$$|a_{ij} - \theta(a_{ij}+b_{ij})| = 0, \tag{15}$$

arranged in the order of descending magnitude: $1 \geqslant \theta_1 \geqslant \theta_2 \geqslant \ldots \geqslant \theta_{l_1} \geqslant 0$. *Then the simultaneous distribution of the θ's, as derived from* (14), *is*

$$\pi^{\frac{1}{2}l_1}\left\{\prod_{i=1}^{l_1} \frac{\Gamma\frac{1}{2}(n-p+l_2+i)}{\Gamma\frac{1}{2}(l_2-l_1+i)\,\Gamma\frac{1}{2}(n-p+i)\,\Gamma\frac{1}{2}i}\right\}$$
$$\times\left\{\prod_{i=1}^{l_1}\theta_i\right\}^{\frac{1}{2}(l_2-l_1-1)}\left\{\prod_{i=1}^{l_1}(1-\theta_i)\right\}^{\frac{1}{2}(n-p-1)}\left\{\prod_{i=1}^{l_1}\prod_{j=i+1}^{l_1}(\theta_i-\theta_j)\right\}\left\{\prod_{i=1}^{l_1} d\theta_i\right\}. \tag{16}$$

Therefore the distributions of U, W and V are those of the functions in (11), (12) and (13), where the θ's follow the distribution (16).

* Lawley (1938). † Hsu (1939), Fisher (1939).

Take the particular case $l_1 = 2$. We have then

$$U = \theta_1\theta_2, \quad W = (1-\theta_1)(1-\theta_2), \quad V = \frac{\theta_1}{1-\theta_1}+\frac{\theta_2}{1-\theta_2},$$

whence $$U^{\frac{1}{2}} + W^{\frac{1}{2}} \leqslant 1, \quad V = W^{-1}(1-U-W),$$

and the distribution (16) becomes

$$\frac{\Gamma(n-p+l_2+1)}{4\Gamma(l_2-1)\Gamma(n-p+1)}(\theta_1\theta_2)^{\frac{1}{2}(l_2-3)}\{(1-\theta_1)(1-\theta_2)\}^{\frac{1}{2}(n-p-1)}(\theta_1-\theta_2)\,d\theta_1 d\theta_2. \tag{17}$$

From (17) we easily obtain the joint distribution of U and W:

$$\frac{\Gamma(n-p+l_2+1)}{4\Gamma(l_2-1)\Gamma(n-p+1)}U^{\frac{1}{2}(l-3)}W^{\frac{1}{2}(n-p-1)}dU\,dW.$$

Integrating with respect to W ranging from 0 to $(1-U^{\frac{1}{2}})^2$, we get the distribution of U:

$$\frac{1}{2B(l_2-1, n-p+2)}U^{\frac{1}{2}(l_2-3)}(1-U^{\frac{1}{2}})^{n-p+1}dU. \tag{18}$$

Similarly, the distribution of W is

$$\frac{1}{2B(l_2, n-p+1)}W^{\frac{1}{2}(n-p-1)}(1-W^{\frac{1}{2}})^{l_2-1}dW. \tag{19}$$

Again, from (18) we find the distribution of V:

$$\frac{\Gamma(n-p+l_2+1)}{4\Gamma(l_2-1)\Gamma(n-p+1)}(1+V)^{-\frac{1}{2}(n-p+3)}\left\{\int_{\frac{4(1+V)}{(2+V)^2}}^{1} y^{\frac{1}{2}(n-p+1)}(1-y)^{\frac{1}{2}(l_2-3)}dy\right\}dV. \tag{20}$$

Now in the particular case considered where $l_1 = 2$ we have either (i) $p = 2$, $l_2 = n_1$ if $p \leqslant n_1$, or (ii) $n_1 = 2$, $l_2 = p$ if $n_1 \leqslant p$. Hence, from (18), (19) and (20) we get the following three pairs of distributions:

$$\frac{1}{2B(n_1-1, n)}U^{\frac{1}{2}(n_1-3)}(1-U^{\frac{1}{2}})^{n-1}dU, \quad 2=p\leqslant n_1,$$

$$\frac{1}{2B(p-1, n-p+2)}U^{\frac{1}{2}(p-3)}(1-U^{\frac{1}{2}})^{n-p+1}dU, \quad 2=n_1\leqslant p,$$

$$\frac{1}{2B(n_1, n-1)}W^{\frac{1}{2}(n-3)}(1-W^{\frac{1}{2}})^{n_1-1}dW, \quad 2=p\leqslant n_1,$$

$$\frac{1}{2B(p, n-p+1)}W^{\frac{1}{2}(n-p-1)}(1-W^{\frac{1}{2}})^{p-1}dW, \quad 2=n_1\leqslant p,$$

$$\frac{\Gamma(n_1+n-1)}{4\Gamma(n_1-1)\Gamma(n-1)}(1+V)^{-\frac{1}{2}(n+1)}dV\int_{4(1+V)(2+V)^{-2}}^{1} y^{\frac{1}{2}(n-1)}(1-y)^{\frac{1}{2}(n_1-3)}dy, \quad 2=p\leqslant n_1,$$

$$\frac{\Gamma(n+1)}{4\Gamma(p-1)\Gamma(n-p+1)}(1+V)^{-\frac{1}{2}(n-p+3)}dV\int_{4(1+V)(2+V)^{-2}}^{1} y^{\frac{1}{2}(n-p+1)}(1-y)^{\frac{1}{2}(p-3)}dy, \quad 2=n_1\leqslant p.$$

The general expression for the moments of U and W can easily be deduced from (16). We have, from (16),

$$C(l_2, n) \int f(\theta; l_2, n) \prod_{i=1}^{l_1} d\theta_i = 1,$$

where
$$C(l_2, n) = \pi^{\frac{1}{2}l_1} \prod_{i=1}^{l_1} \frac{\Gamma\frac{1}{2}(n-p+l_2+i)}{\Gamma\frac{1}{2}(l_2-l_1+i)\,\Gamma\frac{1}{2}(n-p+i)\,\Gamma\frac{1}{2}i},$$

$$f(\theta; l_2, n) = \left\{\prod_{i=1}^{l_1} \theta_i\right\}^{\frac{1}{2}(l_2-l_1-1)} \left\{\prod_{i=1}^{l_1} (1-\theta_i)\right\}^{\frac{1}{2}(n-p-1)} \left\{\prod_{i=1}^{l_1} \prod_{j=i+1}^{l_1} (\theta_i-\theta_j)\right\}.$$

Hence
$$\int f(\theta; l_2, n) \prod_{i=1}^{l_1} d\theta_i = \frac{1}{C(l_2, n)},$$

whence
$$\mathscr{E}(U^{q_1} W^{q_2}) = C(l_2, n) \int f(\theta; l_2+q_1, n+q_2) \prod_{i=1}^{l_1} d\theta_i = \frac{C(l_2, n)}{C(l_2+2q_1, n+2q_2)}$$

$$= \prod_{i=1}^{l_1} \frac{\Gamma\frac{1}{2}(n-p+l_2+i)\,\Gamma\frac{1}{2}(l_2-l_1+2q_1+i)\,\Gamma\frac{1}{2}(n+2q_2-p+i)}{\Gamma\frac{1}{2}(n-p+l_2+2q_1+2q_2+i)\,\Gamma\frac{1}{2}(l_2-l_1+i)\,\Gamma\frac{1}{2}(n-p+i)}.$$

The case of k samples (see p. 221) of two variables ($p=2$) has been studied at length by Pearson & Wilks (1933). The hypothesis H_0 and the functions U and W are called in their paper H_2, U_2 and L_2 respectively. It is found there that the test based on U_2 (i.e. U) cannot be regarded as an adequate test of H_0. We shall show in the next section that H_0 is true if and only if two population constants, called λ_1 and λ_2, both vanish. The disadvantage of U referred to by Pearson and Wilks, may be expressed by saying that it is unlikely to be able to detect the falsehood of H_0 if only one of λ_1 and λ_2 vanishes.

5. *Simplification of the parent distribution function.* The matrix $\|\alpha_{ij}\|$ is positive definite; hence it can be expressed as $\mathbf{CC}'$,* where $\mathbf{C}$ is some non-singular real matrix. Write

$$\mathbf{Y} = \begin{Vmatrix} y_{11} & y_{12} & \cdots & y_{1n_1} \\ y_{21} & y_{22} & \cdots & y_{2n_1} \\ \cdots & \cdots & \cdots & \cdots \\ y_{p1} & y_{p2} & \cdots & y_{pn_1} \end{Vmatrix}, \quad \mathbf{Z} = \begin{Vmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ z_{p1} & z_{p2} & \cdots & z_{pn} \end{Vmatrix}, \quad \mathbf{G} = \begin{Vmatrix} \eta_{11} & \eta_{12} & \cdots & \eta_{1n_1} \\ \eta_{21} & \eta_{22} & \cdots & \eta_{2n_1} \\ \cdots & \cdots & \cdots & \cdots \\ \eta_{p1} & \eta_{p2} & \cdots & \eta_{pn_1} \end{Vmatrix},$$

$$\|\sigma_{ij}\| = \|\alpha_{ij}\|^{-1} = (\mathbf{C}')^{-1}\mathbf{C}^{-1}.$$

THEOREM 3. *Let* $\mathbf{G}$ *be of rank* l *(hence* $l \leqslant p$, $l \leqslant n_1$ *and* l *is the rank of* $GG' = \|\psi_{ij}\|$*), and let* $\lambda_1, \lambda_2, \ldots, \lambda_l$ *be the non-vanishing roots of the determinantal equation* $|\psi_{ij} - \lambda\sigma_{ij}| = 0$. *Let* $\theta_1, \theta_2, \ldots, \theta_{l_1}$ *be the non-identically vanishing roots of the determinantal equation* $|a_{ij} - \theta(a_{ij}+b_{ij})| = 0$. *Then the joint distribution of the* θ*'s, as derived from* (1), *depends upon the parameters* $\lambda_1, \lambda_2, \ldots, \lambda_l$ *alone in such a way that instead of* (1) *we may regard the parent distribution as*

$$(2\pi)^{-\frac{1}{2}p(n_1+n)} \exp(-\tfrac{1}{2}\Psi) \exp\left\{-\frac{1}{2}\sum_{i=1}^{p} a_{ii} - \frac{1}{2}\sum_{i=1}^{p} b_{ii} + \sum_{i=1}^{l} \sqrt{\lambda_i}\, y_{ii}\right\} \Pi\, dy\, dz. \quad (21)$$

* The accent denotes the transposed matrix.

Hence, in studying the distribution of any functions of the θ's, such as U, V and W, we may replace (1) by (21).

LEMMA 1 (Hotelling).* *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be any positive definite matrices of orders* m *and* n *respectively, and* $\mathbf{C}$ *be any real matrix of order* $m \times n$ *and rank* l. *There exist two non-singular real matrices,* $\mathbf{N}_1$ *and* $\mathbf{N}_2$, *such that*

$$\mathbf{N}_1\mathbf{A}\mathbf{N}_1' = \mathbf{I},\dagger \quad \mathbf{N}_2\mathbf{B}\mathbf{N}_2' = \mathbf{I}, \quad \mathbf{N}_1\mathbf{C}\mathbf{N}_2' = \mathbf{D},$$

where $$\mathbf{D} = \left\| \begin{matrix} \mathbf{D}_{\sqrt{\lambda}} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{matrix} \right\|,\dagger \qquad \mathbf{D}_{\sqrt{\lambda}} = \left\| \begin{matrix} \sqrt{\lambda_1} & & & \\ & \sqrt{\lambda_2} & & \\ & & \cdots & \\ & & & \sqrt{\lambda_l} \end{matrix} \right\|,$$

and the λ_i *are the non-vanishing roots of the determinantal equation*

$$|\mathbf{C}\mathbf{B}^{-1}\mathbf{C}' - \lambda\mathbf{A}| = 0.$$

Allowing both $\mathbf{A}$ and $\mathbf{B}$ to be unit matrices in Lemma 1 we get the following:

Corollary. Let $\mathbf{C}$ *be any real matrix of rank* l. *There exist two real orthogonal matrices* $\mathbf{\Gamma}_1$ *and* $\mathbf{\Gamma}_2$ *such that*

$$\mathbf{\Gamma}_1\mathbf{C}\mathbf{\Gamma}_2' = \mathbf{D}$$

where $$\mathbf{D} = \left\| \begin{matrix} \mathbf{D}_{\sqrt{\lambda}} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{matrix} \right\|, \qquad \mathbf{D}_{\sqrt{\lambda}} = \left\| \begin{matrix} \sqrt{\lambda_1} & & & \\ & \sqrt{\lambda_2} & & \\ & & \cdots & \\ & & & \sqrt{\lambda_l} \end{matrix} \right\|,$$

and the λ_i *are the non-vanishing latent roots of the matrix* $\mathbf{C}\mathbf{C}'$.

Proof of Theorem 3. We write $\operatorname{tr}\mathbf{A}$ for the sum of the diagonal elements of any square matrix $\mathbf{A}$. It is easily verified that $\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA})$ whenever $\mathbf{AB}$ is a square matrix.

The expression inside the bracket in (1) may be written as

$$\begin{aligned} &-\tfrac{1}{2}\Psi - \tfrac{1}{2}\operatorname{tr}(\mathbf{CC'YY'}) - \tfrac{1}{2}\operatorname{tr}(\mathbf{CC'ZZ'}) + \operatorname{tr}(\mathbf{CC'GY'}) \\ &\qquad = -\tfrac{1}{2}\Psi - \tfrac{1}{2}\operatorname{tr}(\mathbf{C'YY'C}) - \tfrac{1}{2}\operatorname{tr}(\mathbf{C'ZZ'C}) + \operatorname{tr}(\mathbf{C'GY'C}). \qquad (22) \end{aligned}$$

The λ_i, as defined in Theorem 3, are the non-vanishing latent roots of the equation $|\mathbf{GG'} - \lambda(\mathbf{C'})^{-1}\mathbf{C}^{-1}| = 0$. On pre- and post-multiplying by $|\mathbf{C'}|$ and $|\mathbf{C}|$ respectively this becomes $|\mathbf{C'GG'C} - \lambda\mathbf{I}| = 0$. Hence the λ_i are the non-vanishing latent roots of the matrix $(\mathbf{C'G})(\mathbf{C'G})'$. Therefore, by the corollary to Lemma 1, there exist two real orthogonal matrices, $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$, such that

$$\mathbf{\Gamma}_1\mathbf{C'G}\mathbf{\Gamma}_2 = \mathbf{D} = \left\| \begin{matrix} \mathbf{D}_{\sqrt{\lambda}} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{matrix} \right\|. \qquad (23)$$

* Hotelling (1936), pp. 326–30.

† $\mathbf{I}$ and $\mathbf{O}$ stand for a unit matrix and a zero matrix respectively.

The transformation of variables

$$\mathbf{Y} = (\mathbf{C}')^{-1}\mathbf{\Gamma}_1'\mathbf{U}\mathbf{\Gamma}_2', \quad \mathbf{Z} = (\mathbf{C}')^{-1}\mathbf{\Gamma}_1'\mathbf{V}, \tag{24}$$

where $$\mathbf{U} = \begin{Vmatrix} u_{11} & u_{12} & \cdots & u_{1n_1} \\ u_{21} & u_{22} & \cdots & u_{2n_1} \\ \cdots & \cdots & \cdots & \cdots \\ u_{p1} & u_{p2} & \cdots & u_{pn_1} \end{Vmatrix}, \quad \mathbf{V} = \begin{Vmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ v_{p1} & v_{p2} & \cdots & v_{pn} \end{Vmatrix}$$

are the matrices of the new variables, has a constant Jacobian and leaves the θ's invariant, because the equation (15) becomes

$$| a_{ij}' - \theta(a_{ij}' + b_{ij}') | = 0, \tag{15'}$$

where $$a_{ij}' = \sum_{r=1}^{n_1} u_{ir}u_{jr}, \quad b_{ij}' = \sum_{r=1}^{n} v_{ir}v_{jr}, \quad (i, j = 1, 2, \ldots, p).$$

To show this, we have, remembering the orthogonality of $\mathbf{\Gamma}_2$,

$$\|a_{ij}\| = \mathbf{Y}\mathbf{Y}' = (\mathbf{C}')^{-1}\mathbf{\Gamma}_1'\mathbf{U}\mathbf{U}'\mathbf{\Gamma}_1\mathbf{C}^{-1},$$

$$\|b_{ij}\| = \mathbf{Z}\mathbf{Z}' = (\mathbf{C}')^{-1}\mathbf{\Gamma}_1'\mathbf{V}\mathbf{V}'\mathbf{\Gamma}_1\mathbf{C}^{-1}.$$

Hence the equation (15) is transformed into

$$| \mathbf{U}\mathbf{U}' - \theta(\mathbf{U}\mathbf{U}' + \mathbf{V}\mathbf{V}') | = 0,$$

which is another way of writing (15').

On the other hand, substituting (24) into (22) and remembering (23) we get the expression

$$-\tfrac{1}{2}\Psi - \tfrac{1}{2}\operatorname{tr}(\mathbf{U}\mathbf{U}') - \tfrac{1}{2}\operatorname{tr}(\mathbf{V}\mathbf{V}') + \operatorname{tr}(\mathbf{D}\mathbf{U}')$$
$$= -\tfrac{1}{2}\Psi - \frac{1}{2}\sum_{i=1}^{p} a_{ii}' - \frac{1}{2}\sum_{i=1}^{p} b_{ii}' + \sum_{i=1}^{l} \sqrt{\lambda_i}\, u_{ii}.$$

Replacing again the letters u and v by y and z respectively, we obtain the result.

It may be noticed that

$$\Psi = \operatorname{tr}(\mathbf{C}\mathbf{C}'\mathbf{G}\mathbf{G}') = \operatorname{tr}(\mathbf{C}'\mathbf{G}\mathbf{G}'\mathbf{C}) = \sum_{i=1}^{l} \lambda_i. \tag{25}$$

Remark. For the case of k samples (see Example on p. 221) Fisher (1938) has considered the problem of testing for the colinearity or coplanarity of the k populations. It can be proved that the hypothesis of colinearity (or coplanarity) states that all except one (or two) of the λ's vanish.

6. *Behaviour of V when H_0 is not necessarily true.* The Laplace transform of a probability density function $p(x)$ vanishing identically for $x < 0$, viz. the integral

$$\int_0^\infty e^{-\alpha x} p(x)\, dx,$$

has the following property.

LEMMA 2. *Let $p_n(x)$, $(n=1,2,3,\ldots)$, and $p(x)$ be probability density functions vanishing identically for $x<0$, and such that*

$$\lim_{n\to\infty}\int_0^\infty e^{-\alpha x}p_n(x)\,dx=\int_0^\infty e^{-\alpha x}p(x)\,dx \tag{26}$$

for every $\alpha>0$. Then $\quad \lim_{n\to\infty}\int_0^x p_n(u)\,du=\int_0^x p(u)\,du.$

Proof. The function $f_n(x)=p(x)-p_n(x)$ is summable in $(0,\infty)$ and, by (26),

$$\lim_{n\to\infty}\int_0^\infty e^{-\alpha x}f_n(x)\,dx=0 \quad \text{for every } \alpha>0. \tag{27}$$

The function
$$g_n(z)=\int_0^\infty e^{-zx}f_n(x)\,dx$$

is an analytic function of z, regular at every z with a positive real part and uniformly bounded. By (27) $g_n(z)$ tends to a limit whenever z is real and positive. Hence by Vitali's theorem of convergence* $g_n(z)$ tends to a limit uniformly in the half-plane on the right of the imaginary axis. This limit is an analytic function regular at every z with a positive real part and vanishes whenever z is real and positive; therefore it vanishes identically. Hence

$$\lim_{n\to\infty}\int_0^\infty e^{-\alpha x+itx}p_n(x)\,dx=\int_0^\infty e^{-\alpha x+itx}p(x)\,dx \tag{28}$$

for all $\alpha>0$ and real t. Let

$$L_n(\alpha)=\int_0^\infty e^{-\alpha x}p_n(x)\,dx,\quad L(\alpha)=\int_0^\infty e^{-\alpha x}p(x)\,dx,$$

so that
$$\lim_{n\to\infty}L_n(\alpha)=L(\alpha). \tag{29}$$

It follows from (28) and (29) that

$$\lim_{n\to\infty}\frac{1}{L_n(\alpha)}\int_0^\infty e^{-\alpha u+itu}p_n(u)\,du=\frac{1}{L(\alpha)}\int_0^\infty e^{-\alpha u+itu}p(u)\,du,$$

whence, by a well-known property of the characteristic function,

$$\lim_{n\to\infty}\frac{1}{L_n(\alpha)}\int_0^x e^{-\alpha u}p_n(u)\,du=\frac{1}{L(\alpha)}\int_0^x e^{-\alpha u}p(u)\,du,$$

whence
$$\lim_{n\to\infty}\int_0^x e^{-\alpha u}p_n(u)\,du=\int_0^x e^{-\alpha u}p(u)\,du,$$

whence
$$\lim_{\alpha\to+0}\lim_{n\to\infty}\int_0^x e^{-\alpha u}p_n(u)\,du=\int_0^x p(u)\,du.$$

The order of the above repeated limit can be interchanged because

$$\int_0^x e^{-\alpha u}p_n(u)\,du\to\int_0^x p_n(u)\,du$$

uniformly in n as $\alpha\to+0$. This completes the proof.

* Titchmarsh (1932), p. 168.

We shall now derive an expression for the Laplace integral $\mathscr{E}(\exp(-\alpha V))$. To simplify the notation we shall use a single letter, e.g. y to denote a set of variables, e.g. all the y_{ir}, when they figure as arguments of a function. We shall write dy for $\Pi\, dy$ and a single integration sign for multiple integrals with respect to the variables from $-\infty$ to ∞. We denote the integral $\mathscr{E}(\exp(-\alpha V))$ by $L(\alpha)$.

The following equation can be verified by direct integration, on referring to (10) as the definition of V:

$$\exp(-\tfrac{1}{2}\alpha^2 V) = \int f(y,z,\alpha,t)\,dt,$$

where

$$f(y,z,\alpha,t) = (2\pi)^{-\frac{1}{2}pn_1}\,|\,b_{ij}\,|^{\frac{1}{2}n_1}\exp\left\{-\frac{1}{2}\sum_{i,j=1}^{p} b_{ij}s_{ij} + i\alpha\sum_{i=1}^{p}\sum_{r=1}^{n_1} y_{ir}t_{ir}\right\},$$

$$s_{ij} = \sum_{r=1}^{n_1} t_{ir}t_{jr} \quad (i,j=1,2,\dots,p),$$

and where the letter i, when not figuring as an index, stands for $\sqrt{-1}$. Denoting by $p(y,z)$ the coefficient of $\Pi\, dy\, dz$ in (21), we have

$$L(\tfrac{1}{2}\alpha^2) = \int p(y,z)\,dy\,dz\int f(y,z,\alpha,t)\,dt = \int dt\int p(y,z)f(y,z,\alpha,t)\,dy\,dz, \quad (30)$$

as the change of the order of the integration is obviously legitimate for every real α.

By direct substitution it results

$$\int p(y,z)f(y,z,\alpha,t)\,dy\,dz = (2\pi)^{-\frac{1}{2}p(2n_1+n)}\exp(-\tfrac{1}{2}\Psi)f_1(\alpha,t)f_2(\alpha,t), \quad (31)$$

where
$$f_1(\alpha,t) = \int\exp\left\{-\frac{1}{2}\sum_{i=1}^{p} a_{ii} + \sum_{i=1}^{l}\sqrt{\lambda_i}\,y_{ii} + i\alpha\sum_{i=1}^{p}\sum_{r=1}^{n_1} t_{ir}y_{ir}\right\}dy, \quad (32)$$

$$f_2(\alpha,t) = \int |\,b_{ij}\,|^{\frac{1}{2}n_1}\exp\left\{-\frac{1}{2}\sum_{i=1}^{p} b_{ii} - \frac{1}{2}\sum_{i,j=1}^{p} s_{ij}b_{ij}\right\}dz. \quad (33)$$

The integral (32) is readily evaluated and runs

$$f_1(\alpha,t) = (2\pi)^{\frac{1}{2}pn_1}\exp(\tfrac{1}{2}\Psi)\exp\left\{-\tfrac{1}{2}\alpha^2\sum_{i=1}^{p}\sum_{r=1}^{n_1} t_{ir}^2 - i\alpha\sum_{i=1}^{l}\sqrt{\lambda_i}\,t_{ii}\right\}. \quad (34)$$

The integral (33) can be evaluated by using Wilks's formula for the moments of the generalized variance.* The result is

$$f_2(\alpha,t) = 2^{\frac{1}{2}pn_1}(2\pi)^{\frac{1}{2}pn}K\,|\,\delta_{ij}+s_{ij}\,|^{-\frac{1}{2}(n_1+n)}, \quad (35)$$

where $\delta_{ij} = 1$ if $i=j$ and 0 otherwise, and where

$$K = \prod_{i=1}^{p}\frac{\Gamma\frac{1}{2}(n_1+n-i+1)}{\Gamma\frac{1}{2}(n-i+1)}.$$

* Wilks (1932).

From (34), (35), (31) and (30) we obtain

$$L(\tfrac{1}{2}\alpha^2) = \pi^{-\frac{1}{2}pn_1} K \int |\delta_{ij}+s_{ij}|^{-\frac{1}{2}(n_1+n)} \exp\left\{-\tfrac{1}{2}\alpha^2 \sum_{i=1}^{p} s_{ii} - i\alpha \sum_{i=1}^{l} \sqrt{\lambda_i} t_{ii}\right\} dt,$$

whence

$$L(\alpha) = \pi^{-\frac{1}{2}pn_1} K \int |\delta_{ij}+s_{ij}|^{-\frac{1}{2}(n_1+n)} \exp\left\{-\alpha \sum_{i=1}^{p} s_{ii} - i\sqrt{(2\alpha)} \sum_{i=1}^{l} \sqrt{\lambda_i} t_{ii}\right\} dt. \quad (36)$$

It may be observed that the right-hand side of (36) involves no sum of n terms; hence it is particularly useful when we wish to make n approach infinity.

We may calculate the moments of V by means of (36). Thus

$$\begin{aligned}\mathscr{E}(V) &= \pi^{-\frac{1}{2}pn_1} K \int |\delta_{ij}+s_{ij}|^{-\frac{1}{2}(n_1+n)} \left\{\sum_{i=1}^{p} s_{ii} + \left(\sum_{i=1}^{l} \sqrt{\lambda_i} t_{ii}\right)^2\right\} dt \\ &= \frac{1}{n-p-1}(\Psi + pn_1),\end{aligned} \quad (37)$$

$$\begin{aligned}\mathscr{E}(V^2) &= \pi^{-\frac{1}{2}pn_1} K \int |\delta_{ij}+s_{ij}|^{-\frac{1}{2}(n_1+n)} \left\{\left(\sum_{i=1}^{p} s_{ii}\right)^2 + 2\left(\sum_{i=1}^{p} s_{ii}\right)\left(\sum_{i=1}^{l} \sqrt{\lambda_i}\, t_{ii}\right)^2 \right. \\ &\qquad \left. + \frac{1}{3}\left(\sum_{i=1}^{l} \sqrt{\lambda_i} t_{ii}\right)^4\right\} dt \\ &= \frac{\Psi^2}{(n-p-1)(n-p-3)} - \frac{4\sum_{i \neq j} \lambda_i \lambda_j}{(n-p)(n-p-1)(n-p-3)} \\ &\qquad + 2\Psi\left\{\frac{pn_1+2}{(n-p-1)(n-p-3)} - \frac{2(p-1)(n_1-1)}{(n-p)(n-p-1)(n-p-3)}\right\} \\ &\qquad + \frac{pn_1(pn_1+2)}{(n-p-1)(n-p-3)} - \frac{2(p-1)(n_1-1)}{(n-p)(n-p-1)(n-p-3)}.\end{aligned} \quad (38)$$

Remark. Consider again the case of k samples (see Example on p. 221) and write V_k and Ψ_k correspondingly. It is seen from (37) that the statistic

$$(M-k-p-1)V_k - p(k-1)$$

is an unbiassed estimate of Ψ_k. We may regard V_k as a generalization of V_2, which is the "Studentized" D^2 statistic* but for a constant factor. V_2 is, of course, also identical with Hotelling's generalized "Student's" ratio except for a constant factor.

We shall next study the behaviour of V as $n \to \infty$. As n is now allowed to vary, we shall attach to various letters the index n. In particular we write Ψ_n for Ψ to emphasize the fact that the value of Ψ depends also on n, a fact which is not brought out by the formal definition of Ψ. This is because the η_{ir}, as we explained on p. 221, are themselves linear functions of the original population constants with coefficients depending upon n.

* Bose and Roy (1938).

It is in general true that, except when H_0 is true, in which case $\Psi_n \equiv 0$, Ψ_n is either $O(1)$ or $O(n)$ as $n \to \infty$. That both are possible is seen in the Example on p. 221 for $k = 2$; Ψ_n is $O(n)$ or $O(1)$ according as $m_1 = m_2 \to \infty$ or one of m_1, m_2 remains fixed while the other tends to infinity.

THEOREM 4. *If $\Psi_n = O(n)$, then the mean of V is $\frac{1}{n}\Psi_n + O\left(\frac{1}{n}\right)$ and the variance of V is $O\left(\frac{1}{n}\right)$. Hence the random variable $V - \frac{1}{n}\Psi_n \to 0$ in probability.*

This is an immediate consequence of (37) and (38).

If $\Psi_n = O(1)$, and if Ψ_n tends to a limit as $n \to \infty$, the limiting distribution of nV will be that of S (cf. (2)) with Ψ replaced by its limit. This is our next theorem. We assume now that

$$\lim_{n\to\infty} \Psi_n = \Psi_0. \tag{39}$$

The case $\Psi_n \equiv 0$ (i.e. H_0 is true) is covered by (39) on putting $\Psi_0 = 0$. But (39) may also be regarded as covering the case $\Psi_n = O(n)$ in the sense that the repeated sampling is not made from the same population, but from populations whose distribution constants vary with n in such a way that (39) holds true. Thus, for example, in the case of two samples with $m_1 = m_2$ (see p. 222) we have

$$\Psi_n = \tfrac{1}{2}m\Phi = \tfrac{1}{4}(n+2)\Phi,$$

where

$$\Phi = \sum_{i,j=1}^{p} \alpha_{ij}(\xi_{i1}-\xi_{i2})(\xi_{j1}-\xi_{j2}).$$

Here we have $\Psi_n = O(n)$ if the population constant $\Phi > 0$. Ψ_n will be $O(1)$ if we consider Φ as decreasing to the order n^{-1} as n increases indefinitely. This idea of regarding population constants as varying with the sample size can be found in the works of Fisher* and Neyman.†

THEOREM 5. *If Ψ_n tends to a limit Ψ_0 as $n \to \infty$, then*

$$\lim_{n\to\infty} Pr\{nV \leqslant x\} = \int_0^x p(x)\,dx, \tag{40}$$

where

$$p(x) = 2^{-\frac{1}{2}pn_1} x^{\frac{1}{2}pn_1-1} e^{-\frac{1}{2}(x+\Psi_0)} \sum_{h=0}^{\infty} \frac{\Psi_0^h x^h}{4^h h!\,\Gamma(h+\frac{1}{2}pn_1)}. \tag{41}$$

In particular, if $\Psi_0 = 0$ (i.e. H_0 is true), the limiting distribution of nV is that of χ^2 with pn_1 degrees of freedom.

Proof. Write $L_1^{(n)}(\alpha)$ for the Laplace transform of the probability density function of nV, so that

$$L_1^{(n)}(\alpha) = \mathscr{E}(\exp(-\alpha nV)) = L(n\alpha).$$

* Fisher (1928), p. 663.

† Neyman (1937), pp. 169–70, (1938), pp. 70–1.

If we replace α by $n\alpha$ in (36) and subsequently make the change of variables $t_{ir} = n^{-\frac{1}{2}}\tau_{ir}$, we get

$$L_1^{(n)}(\alpha) = \pi^{-\frac{1}{2}pn_1} n^{-\frac{1}{2}pn_1} K_n \int \left| \delta_{ij} + \frac{s'_{ij}}{n} \right|^{-\frac{1}{2}(n_1+n)} \exp\left\{ -\alpha \sum_{i=1}^{p} s'_{ii} - i\sqrt{(2\alpha)} \sum_{i=1}^{l} \sqrt{\lambda_i}\,\tau_{ii} \right\} d\tau, \tag{42}$$

where K_n is now written for K and

$$s'_{ij} = \sum_{r=1}^{n_1} \tau_{ir}\tau_{jr}, \quad (i, j = 1, 2, \ldots, p).$$

We shall now find the limit of $L_1^{(n)}(\alpha)$ as $n \to \infty$. We have, using Stirling's formula,

$$\lim_{n\to\infty} n^{-\frac{1}{2}pn_1} K_n = 2^{-\frac{1}{2}pn_1}. \tag{43}$$

For every system of fixed values of the τ_{ir} we have

$$\left| \delta_{ij} + \frac{s'_{ij}}{n} \right| = 1 + \frac{1}{n} \sum_{i=1}^{p} s'_{ii} + O\left(\frac{1}{n^2}\right),$$

whence

$$\lim_{n\to\infty} \left| \delta_{ij} + \frac{s'_{ij}}{n} \right|^{\frac{1}{2}(n_1+n)} = \exp\left(-\frac{1}{2} \sum_{i=1}^{p} s'_{ii} \right). \tag{44}$$

Calling I_n the integral in (42), we have

$$\begin{aligned} I_n = \int \left\{ \left| \delta_{ij} + \frac{s'_{ij}}{n} \right|^{-\frac{1}{2}(n_1+n)} - \exp\left(-\tfrac{1}{2}\textstyle\sum s'_{ii}\right) \right\} \exp\left\{ -\alpha \sum_{i=1}^{p} s'_{ii} - i\sqrt{(2\alpha)} \sum_{i=1}^{l} \sqrt{\lambda_i}\,\tau_{ii} \right\} d\tau \\ + \int \exp\left\{ -(\tfrac{1}{2}+\alpha) \sum_{i=1}^{p} s_{ii} - i\sqrt{(2\alpha)} \sum_{i=1}^{l} \sqrt{\lambda_i}\,\tau_{ii} \right\} d\tau \\ = A_n + B_n. \end{aligned} \tag{45}$$

The absolute value of the integrand of A_n is less than

$$2\exp\left(-\alpha \sum_{i=1}^{p} s'_{ii} \right).$$

Hence by (44) and dominated convergence, $A_n \to 0$ as $n \to \infty$. The value of B_n is easily found to be

$$B_n = (2\pi)^{\frac{1}{2}pn_1} (1+2\alpha)^{-\frac{1}{2}pn_1} \exp\left(-\frac{\alpha\Psi_n}{1+2\alpha} \right).$$

From (42), (43) and (45) we obtain the result

$$\lim_{n\to\infty} L_1^{(n)}(\alpha) = (1+2\alpha)^{-\frac{1}{2}pn_1} \exp\left(-\frac{\alpha\Psi_0}{1+2\alpha} \right). \tag{46}$$

If $p(x)$ is defined as in (41), then the integral

$$\int_0^\infty e^{-\alpha x} p(x)\, dx$$

is identically equal to the right-hand side of (46). Theorem 5 is thus proved on remembering Lemma 2.

7. *Behaviour of W when H_0 is not necessarily true and n is large.* We shall establish the following theorem.

THEOREM 6. *If $\Psi_n = O(1)$, the statistics nV and $-n\log W$ tend to be certainly identical in the sense that*

$$\lim_{n\to\infty}(nV+n\log W)=0 \textit{ in probability.} \tag{47}$$

Corollary. If Ψ_n tends to a limit Ψ_0 as $n\to\infty$, $-n\log W$ has the same limiting distribution (40) *as nV as $n\to\infty$.*

Remark. Besides the Corollary there is another significance to be attached to Theorem 6. Since the test functions V and W are not functionally related (except when $p=1$ or $n_1=1$), there is the question of choosing one of them to be consistently used in carrying out the actual tests. This can only be decided by a comparison of their power* to detect the falsehood of H_0 when the η_{ir} do not all vanish. While this may be a difficult problem for small samples, Theorem 6 appears to have answered the question for large samples. In fact, if n is large, it is almost certain that the values of nV and $-n\log W$ calculated from the sample will differ very little. That $-n\log W$ and nV tend to have the same power function is a consequence of the Corollary.

Proof of Theorem 6. For every θ such that $0\leqslant\theta\leqslant 1$ we have

$$0\leqslant\frac{\theta}{1-\theta}+\log(1-\theta)\leqslant\frac{1}{2}\left(\frac{\theta}{1-\theta}\right)^2.$$

Hence, remembering (12) and (13),

$$0\leqslant V+\log W\leqslant\frac{1}{2}\sum_{i=1}^{l_1}\left(\frac{\theta_i}{1-\theta_i}\right)^2\leqslant\frac{1}{2}\left(\sum_{i=1}^{l_1}\frac{\theta_i}{1-\theta_i}\right)^2=\tfrac{1}{2}V^2,$$

whence, for every $\eta>0$,

$$\begin{aligned}Pr\{|\,nV+n\log W\,|>\eta\}&=Pr\{nV+n\log W>\eta\}\leqslant Pr\left\{\frac{n}{2}V^2>\eta\right\}\\&\leqslant\frac{1}{\eta}\mathscr{E}\left(\frac{n}{2}V^2\right)=O\left(\frac{1}{n}\right)\end{aligned}$$

because of (38). This establishes (47).

Proof of the Corollary. Let

$$F_n(x)=Pr\{nV\leqslant x\},\quad G_n(x)=Pr\{-n\log W\leqslant x\}.$$

Then for every x and $\eta>0$ we have†

$$0\leqslant G_n(x)-F_n(x)\leqslant F_n(x+\eta)-F_n(x-\eta)+Pr\{nV+n\log W\geqslant\eta\}.$$

* Neyman & Pearson (1936, 1938).

† Fréchet (1937), p. 164.

Let $F(x)$ be the limit of $F_n(x)$ as $n \to \infty$, which is known to exist by virtue of Theorem 6. We have then

$$0 \leqslant G_n(x) - F_n(x) \leqslant | F_n(x+\eta) - F(x+\eta) | + | F_n(x-\eta) - F(x-\eta) | \\ + F(x+\eta) - F(x-\eta) + Pr\{nV + n \log W \geqslant \eta\}. \tag{48}$$

Given any $\epsilon > 0$, choose and fix an $\eta > 0$ so small that $F(x+\eta) - F(x-\eta) < \epsilon$. By (47) the rest of the terms in the right-hand side of (48) is smaller than ϵ for all sufficiently large n. Hence $G_n(x) - F_n(x) \to 0$, i.e. $G_n(x) \to F(x)$, which completes the proof.

Summary

The Wilks-Lawley hypothesis concerning population means of multivariate normal populations is put in the canonical form H_0 (§ 1), and, assuming the population variances and covariances known, the test function S is derived together with its exact distribution (§ 2). In the case where the population variances and covariances are unknown, two possible test functions, denoted by V and W are considered (§§ 3 and 4), and their distributions in certain special cases are given. In § 5 it is shown that the sample space can be so transformed that all the variables are independently distributed and that only a minimum number of unknown parameters remain. These parameters are the roots of a certain determinantal equation; the hypothesis H_0, and the hypotheses of colinearity and coplanarity of populations, all specify the values zero for all or some of these parameters. Returning to the test functions V and W, it is shown that as the sample size n increases indefinitely the two functions nV and $-n \log W$ tend to be certainly identical ((47), § 7), and both of them tend to have the same distribution function as S ((40), § 6).

REFERENCES

Bose, R. C. (1936). "On the exact distribution of the D^2-statistic." *Sankhya*, **2**, 143–54.

Bose, R. C. & Roy, S. N. (1938). "The distribution of the 'Studentized' D^2-statistic." *Proceedings of the first session of the Indian Statistical Conference, Calcutta*, pp. 19–38.

Fisher, R. A. (1928). "The general sampling distribution of the multiple correlation coefficient." *Proc. Roy. Soc.* **A**, **121**, 653–73.

—— (1938). "The statistical utilization of multiple measurements." *Ann. Eugen., Lond.*, **8**, 376–86.

—— (1939). "The sampling distribution of some statistics obtained from non-linear equations." *Ann. Eugen., Lond.*, **9**, 238.

Fréchet, M. (1937). *Généralités sur les probabilités. Variables aléatoires.* Paris.

Hotelling, H. (1931). "The generalization of Student's ratio." *Ann. Math. Statist.* **2**, 359–78.

—— (1936). "Relations between two sets of variates." *Biometrika*, **28**, 321–77.

Hsu, P. L. (1938). "Notes on Hotelling's generalized T." *Ann. Math. Statist.* **9**, 231–43.

—— (1939). "On the distribution of roots of certain determinantal equations." *Ann. Eugen., Lond.*, **9**, 250.

Kolodziezczyk, St (1935). "On an important class of statistical hypotheses." *Biometrika*, **27**, 161–90.

Lawley, D. N. (1938). "A generalization of Fisher's z-test." *Biometrika*, **30**, 180–8.

Mahalanobis, P. C. (1936). "On the generalized distance in statistics." *Proc. Nat. Inst Sci. India*, **2**, 49–55.

Neyman, J. (1937). "Smooth test for goodness of fit." *Skand. AktuarTidskr.* pp. 149–99.

—— (1938). "Tests of statistical hypotheses which are unbiassed in the limit." *Ann. Math. Statist.* **9**, 69–86.

Neyman, J. & Pearson, E. S. (1928). "On the use and interpretation of certain test criteria for purposes of statistical inference." *Biometrika*, **20** A, 175–240.

—— —— (1936, 1938). "Contributions to the theory of testing statistical hypotheses. I." *Statist. Res. Mem.* **1**, 1–37. "Contributions to the theory of testing statistical hypotheses. II." *Statist. Res. Mem.* **2**, 25–57.

Pearson, E. S. & Wilks, S. S. (1933). "Methods of statistical analysis appropriate for k samples of two variables." *Biometrika*, **25**, 353–78.

Tang, P. C. (1938). "The power function of the analysis of variance tests." *Statist. Res. Mem.* **2**, 126–49.

Titchmarsh, E. C. (1932). *The Theory of Functions*. Oxford University Press.

Wilks, S. S. (1932). "Certain generalization in the analysis of variance." *Biometrika*, **24**, 471–94.

THE DERIVATION OF THE FIFTH AND SIXTH MOMENTS OF THE DISTRIBUTION OF b_2 IN SAMPLES FROM A NORMAL POPULATION

By C. T. HSU and D. N. LAWLEY

Department of Statistics, University College, London

1. Introduction

The method of combinatorial analysis was first introduced in 1928 in a paper by R. A. Fisher (1929). In this paper Fisher defined new symmetric functions $k_1, k_2, k_3, \ldots$ of the observations for samples of a given size, and gave simple rules for determining the cumulants or semi-invariants of their joint sampling distribution. This method is especially valuable for the case where the sampled population is normal, and in this case its use reduces considerably the labour of deriving the higher sampling moments of the distribution of product moment statistics. Two further papers appeared later, one a joint paper of R. A. Fisher and J. Wishart (1931) in which further rules were given, the other by Wishart (1930) in which he described applications of the theory and gave a list of higher order formulae for the normal case. By means of these formulae E. S. Pearson (1930) was able to derive the first four moments of the sampling distribution of the statistics $\sqrt{b_1}$ and b_2 for the case when the sampled population was assumed to be normal.

The object of the present paper is to derive formulae for $\kappa(4^5)$ and $\kappa(4^6)$ (quantities which are defined below) and then to use these to determine the fifth and sixth moment coefficients of the distribution of b_2. This work is preliminary to further investigations regarding the sampling distribution of this expression.

If $x_1, x_2, \ldots, x_n$ be a sample of size n from a given population, then adopting Fisher's definition of the symmetric functions k_1, k_2, k_3, k_4, we have

$$k_1 = m_1,$$

$$k_2 = \frac{n}{n-1} m_2,$$

$$k_3 = \frac{n^2}{(n-1)(n-2)} m_3,$$

$$k_4 = \frac{n^2}{(n-1)(n-2)(n-3)} \{(n+1)m_4 - 3(n-1)m_2^2\},$$

where
$$m_r = \frac{1}{n} \overset{n}{\underset{1}{S}} (x-\bar{x})^r, \quad \bar{x} = \frac{1}{n} \overset{n}{\underset{1}{S}} x.$$

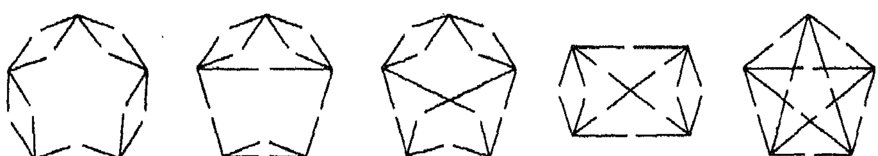
The coefficients k_r are chosen so that $E(k_r) = \kappa_r$, where κ_r is the rth cumulant of the given population, and the pth cumulant of the distribution of k_r is as usual denoted by $\kappa(r^p)$.

$\kappa(r^p)$ is expressible as the sum of a number of terms of order pr consisting of powers and products of the coefficients $\kappa_2, \kappa_3, \ldots, \kappa_{pr}$, and to each such term there corresponds a number of two-way partitions whose coefficients have to be evaluated. As in this paper we are, however, assuming the sampled population to be normal, the only non-vanishing cumulant is κ_2, which is equal to σ^2, the variance of the distribution. The only term, therefore, which has to be evaluated is the one containing a power of κ_2.

2. THE DERIVATION OF $\kappa(4^5)$

For a full explanation of what follows the reader is referred to the three papers by Fisher and Wishart already cited.

To determine the formula for $\kappa(4^5)$ we must find the coefficients of all the two-way patterns for the term in κ_2^{10}. There are altogether five such patterns, which are given in Wishart's paper, and they are reproduced below. In each pattern it will be noted that there are five corners with four arms attached, each corner representing a k_4, and ten connections between the pairs of arms, each connection representing a κ_2.



A B C D E

Fig. 1.

For each of these five patterns, we must determine (1) the numerical coefficient, (2) the n-coefficient.

It is not necessary to give in detail all the working required for doing this, but we shall give two examples for each process and also a summary of the results.

(1) The numerical coefficient is obtained by enumerating all the ways in which the pattern can be connected up, regarding as a separate entity each corner and also each arm attached to that corner.

(*a*) Consider first the pattern A.

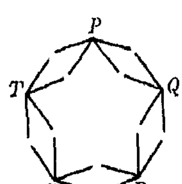


Fig. 2.

There are $\frac{1}{2} \times 4! = 12$ ways of determining the cyclic order of the five corners.

There are six ways of choosing a pair of arms belonging to corner P to connect up to Q, and similarly for the other four corners.

Finally, there are two ways of joining up the double arms between each pair of adjacent corners. The numerical coefficient of A is therefore $12 \times 6^5 \times 2^5 = 12 \times 12^5$.

(*b*) Consider the pattern D.

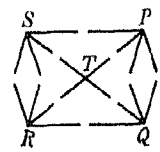


Fig. 3.

There are five ways of choosing which corner is to form the centre T.

There are three ways of choosing the pairs of corners P, Q and R, S which are linked up by double arms, and two ways of connecting up these two pairs.

The arms of corner T may be arranged in $4! = 24$ ways, while those of each of the other corners may be arranged in 12 ways.

Finally, there are two ways of connecting up each of the two double arms.

Hence the numerical coefficient* of D is

$$5 \times (3 \times 2) \times (24 \times 12^4) \times 2^2 = 240 \times 12^5.$$

We may remark that the five normal patterns for $\kappa(4^5)$ can all be obtained from those of $\kappa(4^4)$ by the insertion of a fifth corner. This fact allows us to derive the numerical coefficients by a different method, which may be used as a check. This method must, however, be employed with care as otherwise it is liable to suggest a wrong result. We shall illustrate it by considering again the pattern A.

The pattern A may be formed from the smaller pattern which has a numerical coefficient of 62208, by breaking one of the double arms and inserting a new four-way corner.

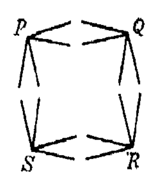


Fig. 4.

The break may be made in four places and the arms of the new corner may then be joined up in 24 ways. The numerical coefficient of A would thus at first sight appear to be $62208 \times 4 \times 24 = 24 \times 12^5$.

It must, however, be remembered that when for instance the break in the smaller pattern is made between P and Q, the broken pattern corresponds to two arrangements of the unbroken pattern, since there are two ways of joining up the double arms between P and Q. Hence the true value of the numerical coefficient of A is in fact half the above number, i.e. 12×12^5.

* There is apparently a mistake in Wishart's paper, the numerical coefficient of D being given there as 120×12^5. This will of course mean that the approximate value of $\kappa(4^5)$ found in that paper is incorrect.

(2) The n-coefficient.

In a previous paper (Fisher & Wishart, 1931) rules were given for deriving the n-coefficients of a given pattern from those of non-vanishing patterns of a lower order which are already known. Thus we can derive the n-coefficient of A, which is one of the patterns of $\kappa(4^5)$, from those of two of the patterns of $\kappa(4^4)$. This is illustrated in Fig. 5, the Greek letters below each pattern are used to denote the n-coefficient of that pattern.

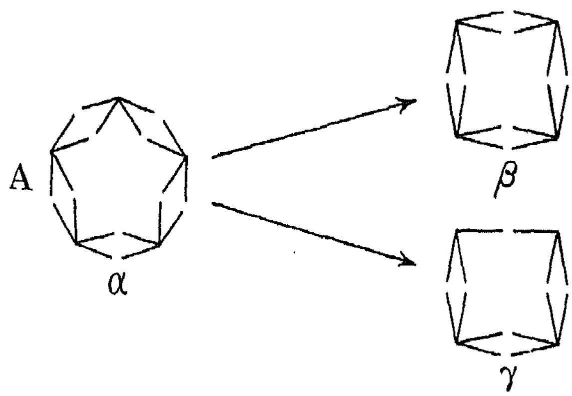


Fig. 5.

Using the rule for adding a new k_4, we have

$$\alpha = 2\frac{(-1)}{(n-2)(n-3)}\beta + \frac{n(n+1)}{(n-1)(n-2)(n-3)}\gamma,$$

and we also know that

$$\beta = \frac{n(n+1)\,n^4-8n^3+21n^2-14n+4}{(n-1)^3(n-2)^3(n-3)^3}$$

and

$$\gamma = \frac{n^3}{(n-1)^3(n-2)^3}.$$

Hence we obtain the result

$$\alpha = \frac{n(n+1)}{(n-1)^4(n-2)^4(n-3)^4}\{n^6-11n^5+45n^4-85n^3+70n^2-36n+8\}.$$

This is the n-coefficient of A.

Similarly for pattern B.

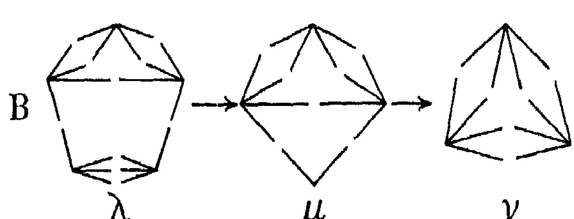


Fig. 6

Here we have $$\lambda = \frac{n+1}{(n-1)(n-2)(n-3)}\mu,$$

$$\mu = \frac{n}{n-1}\nu$$

and $$\nu = \frac{n(n+1)(n^2-5n+2)}{(n-1)^4(n-2)^3(n-1)^3}.$$

Hence $$\lambda = \frac{n^2(n+1)^2(n^2-5n+2)}{(n-1)^4(n-2)^3(n-3)^3}.$$

This is the n-coefficient of B.

The complete results are as follows:

Pattern	Numerical coefficient/12^5	n-coefficient $\div n(n+1)/\{(n-1)^4(n-2)^4(n-3)^4\}$
A	12	$n^6-11n^5+45n^4-85n^3+70n^2-36n+8$
B	80	$n^6-9n^5+23n^4-7n^3-28n^2+12n$
C	120	$n^6-12n^5+52n^4-94n^3+59n^2-30n+8$
D	240	$n^6-13n^5+59n^4-103n^3+48n^2-24n+8$
E	32	$n^6-15n^5+80n^4-160n^3+75n^2-41n+12$

The formula for $\kappa(4^5)$ is calculated by finding the sum of products of the numerical coefficients with the n-coefficients. We thus obtain the result

$$\kappa(4^5) = \frac{4\times 12^5 n(n+1)}{(n-1)^4(n-2)^4(n-3)^4}\{121n^6-1473n^5+6335n^4 -10{,}675n^3+4900n^2-2536n+840\}\,\kappa_2^{10}.$$

3. The derivation of $\kappa(4^6)$

There are altogether 17 normal patterns for $\kappa(4^6)$, which are shown in Fig. 7.

All the patterns except the last two (R and S) may be obtained from those of $\kappa(4^5)$ by the insertion of a sixth corner, while the patterns R and S may be obtained from those of $\kappa(4^4)$ by the insertion of a pair of corners. The numerical coefficients may as before be calculated in two ways and we shall content ourselves with a summary of the results. We shall, however, give two examples of the derivation of the n-coefficients. That for the pattern A is shown in Fig. 8. Here we have

$$\alpha = 2\times\frac{(-1)}{(n-2)(n-3)}\beta + \frac{n(n+1)}{(n-1)(n-2)(n-3)}\gamma,$$

$$\beta = \frac{n(n+1)}{(n-1)^4(n-2)^4(n-3)^4}\{n^6-11n^5+45n^4-85n^3+70n^2-36n+8\},$$

$$\gamma = \frac{n}{(n-1)(n-2)}\delta = \frac{n}{(n-1)(n-2)}\times\frac{n^3}{(n-1)^3(n-2)^3}$$

$$= \frac{n^4}{(n-1)^4(n-2)^4}.$$

Hence $$\alpha = \frac{n(n+1)}{(n-1)^5(n-2)^5(n-3)^5}\{n^8-14n^7+78n^6-220n^5+341n^4 -310n^3+212n^2-88n+16\}.$$

This is the n-coefficient of A.

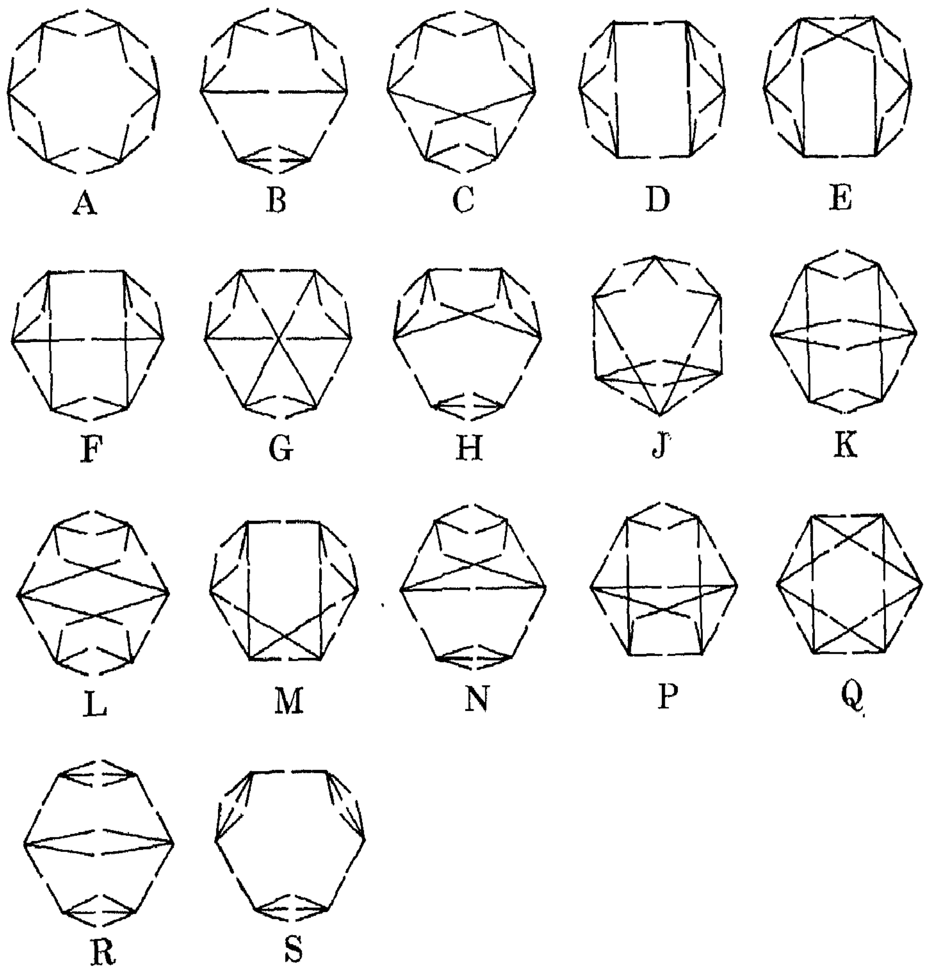


Fig. 7.

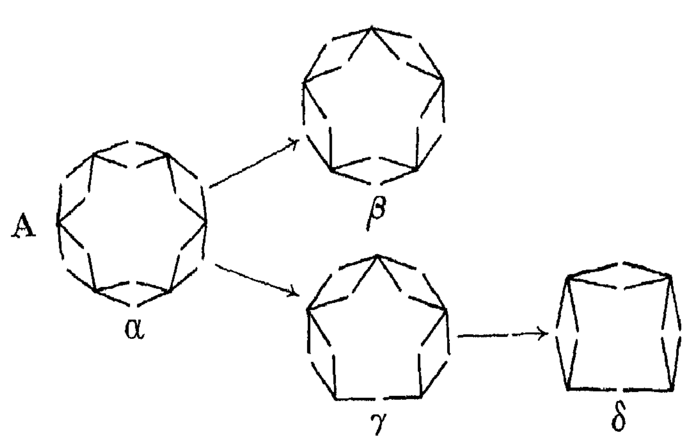


Fig. 8.

Similarly for the pattern S.

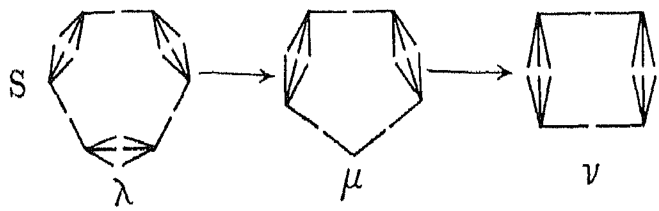


Fig. 9.

$$\lambda = \frac{(n+1)}{(n-1)(n-2)(n-3)}\mu,$$

$$\mu = \frac{n}{n-1}\nu,$$

and
$$\nu = \frac{n^2(n+1)^2}{(n-1)^3(n-2)^2(n-3)^2}.$$

Hence the n-coefficient of S is

$$\lambda = \frac{n^3(n+1)^3}{(n-1)^5(n-2)^3(n-3)^3}.$$

The complete results are given in the following table. Here c denotes the numerical coefficient divided by k, where $k = 2^{14} \times 3^4 \times 5 = 6{,}635{,}520$ and a_r is the coefficient of n^r in the expression for n-coefficient divided by N, where

$$N = \frac{n(n+1)}{(n-1)^5(n-2)^5(n-3)^5}.$$

Table of coefficients

Pattern	c	a_8	a_7	a_6	a_5	a_4	a_3	a_2	a_1	a_0
A	27	1	−14	78	−220	341	−310	212	−88	16
B	216	1	−12	54	−100	33	96	−80	24	0
C	324	1	−15	88	−254	373	−279	178	−76	16
D	324	1	−14	70	−140	53	138	−116	24	0
E	162	1	−15	90	−266	389	−267	160	−176	16
F	648	1	−16	99	−292	403	−236	153	−64	16
G	216	1	−16	104	−324	449	−208	110	−52	16
H	432	1	−13	60	−106	17	111	−62	24	0
J	1296	1	−16	100	−300	421	−236	126	−64	16
K	216	1	−16	103	−320	451	−220	101	−52	16
L	324	1	−16	99	−292	403	−236	153	−64	16
M	2592	1	−17	111	−338	451	−193	101	−52	16
N	432	1	−13	60	−106	17	111	−62	24	0
P	1296	1	−18	129	−442	675	−342	195	−94	24
Q	432	1	−19	144	−521	825	−385	210	−107	28
R	72	1	−8	18	4	−47	12	36	0	0
S	16	1	−8	18	4	−47	12	36	0	0

The formula for $\kappa(4^6)$ derived from these results is

$$\kappa(4^6) = 6,635,520\frac{n(n+1)}{(n-1)^5(n-2)^5(n-3)^5}\{9025n^8-145,532n^7+920,610n^6 -2,775,248n^5+3,759,853n^4-1,717,116n^3+946,872n^2 -476,064n+136,080\}\kappa_2^{12}.$$

4. THE MOMENTS OF THE DISTRIBUTION OF b_2

The formulae of $\kappa(4^2)$, $\kappa(4^3)$ and $\kappa(4^4)$ have been found previously and we will give the results:

$$\kappa(4^2) = \frac{24n(n+1)}{(n-1)(n-2)(n-3)}\kappa_2^4,$$

$$\kappa(4^3) = \frac{1728n(n+1)(n^2-5n+2)}{(n-1)^2(n-2)^2(n-3)^2}\kappa_2^6,$$

$$\kappa(4^4) = \frac{6912n(n+1)(53n^4-428n^3+1025n^2-474n+180)}{(n-1)^3(n-2)^3(n-3)^3}\kappa_2^8.$$

We may now calculate the moments of the distribution of k_4 as follows:

$$\mu(4^2) = \kappa(4^2),$$

$$\mu(4^3) = \kappa(4^3),$$

$$\begin{aligned}\mu(4^4) &= \kappa(4^4)+3\kappa^2(4^2)\\ &= \frac{1728n(n+1)}{(n-1)^3(n-2)^3(n-3)^3}\{n^5+207n^4-1707n^3+4105n^2-1902n+720\}\kappa_2^8,\end{aligned}$$

$$\begin{aligned}\mu(4^5) &= \kappa(4^5)+10\kappa(4^2)\kappa(4^3)\\ &= \frac{12^4.4n(n+1)}{(n-1)^4(n-2)^4(n-3)^4}\{5n^7+1402n^6-17,516n^5+75,870n^4\\ &\qquad -128,205n^3+59,000n^2-30,492n+10,080\}\kappa_2^{10},\end{aligned}$$

$$\begin{aligned}\mu(4^6) &= \kappa(4^6)+15\kappa(4^4)\kappa(4^2)+10\kappa^2(4^3)+15\kappa^3(4^2)\\ &= \frac{12^4.10n(n+1)}{(n-1)^5(n-2)^5(n-3)^5}\{n^{10}+770n^9+278,359n^8-4,603,808n^7\\ &\qquad +29,339,555n^6-88,717,430n^5+120,300,577n^4-55,075,788n^3\\ &\qquad +30,365,028n^2-15,250,464n+4,354,560\}\kappa_2^{12}.\end{aligned}$$

The quantity b_2 is defined to be m_4/m_2^2 and, using the definitions of k_2 and k_4 given in § 1, we see that

$$\begin{aligned}b_2 &= \frac{(n-1)(n-2)(n-3)}{n^2(n+1)}\frac{k_4}{m_2^2}+\frac{3(n-1)}{n+1}\\ &= \frac{(n-2)(n-3)k_4}{(n-1)(n+1)k_2^2}+\frac{3(n-1)}{n+1}.\end{aligned}$$

Thus the average value of b_2 is $3(n-1)/(n+1)$ and, for $r > 1$, the rth moment of the distribution of b_2 is given by

$$\mu_r(b_2) = \left\{\frac{(n-2)(n-3)}{(n-1)(n+1)}\right\}^r \mu_r(k_4 k_2^{-2})$$

$$= \left\{\frac{(n-2)(n-3)}{(n-1)(n+1)}\right\}^r \mu(4^r 2^{-r}).$$

Now Fisher (1930) has shown that

$$\mu(\ldots 5^c 4^b 3^a) = \mu(\ldots 5^c 4^b 3^a 2^{-r}) \frac{(n-1)\ldots(n+2r-3)}{(n-1)^r} \kappa_2^r,$$

and in particular

$$\mu(4^r) = \mu(4^r 2^{-2r}) \frac{(n-1)(n+1)\ldots(n+4r-3)}{(n-1)^{2r}} \kappa_2^{2r}.$$

Thus
$$\mu_r(b_2) = \frac{(n-1)^r (n-2)^r (n-3)^r}{(n+1)^r (n-1)(n+1)\ldots(n+4r-3)} \frac{\mu(4^r)}{\kappa_2^{2r}}.$$

Hence, finally, we obtain the following results:

$$\mu_2(b_2) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)},$$

$$\mu_3(b_2) = \frac{1728n(n-2)(n-3)(n^2-5n+2)}{(n+1)^3(n+3)(n+5)(n+7)(n+9)},$$

$$\mu_4(b_2) = \frac{1728n(n-2)(n-3)(n^5+207n^4-1707n^3+4105n^2-1902n+720)}{(n+1)^4(n+3)(n+5)(n+7)(n+9)(n+11)(n+13)},$$

$$\mu_5(b_2) = \frac{12^4 . 4n(n-2)(n-3)}{(n+1)^5(n+3)(n+5)\ldots(n+17)}\{5n^7+1402n^6-17{,}516n^5+75{,}870n^4$$
$$-128{,}205n^3+59{,}000n^2-30{,}492n+10{,}080\},$$

$$\mu_6(b_2) = \frac{12^4 . 10n(n-2)(n-3)}{(n+1)^6(n+3)(n+5)\ldots(n+21)}\{n^{10}+770n^9+278{,}359n^8-4{,}603{,}808n^7$$
$$+29{,}339{,}555n^6-88{,}717{,}430n^5+120{,}380{,}577n^4-55{,}075{,}788n^3$$
$$+30{,}365{,}028n^2-15{,}250{,}464n+4{,}354{,}560\}.$$

5. Check with McKay's results in the case $n = 4$

Since the formulae which we have obtained are somewhat complicated, it is desirable to provide some sort of check on them. This is supplied by using the results obtained by McKay (1933), who found in an exact form the distribution of b_2 for $n = 4$, i.e. for samples of size 4.

Putting $x = b_2$, the distribution function $f(x)$, of x, is given by

$$f(x) = \frac{3}{2^{\frac{5}{2}}(7-3x)^{\frac{1}{2}}} F\left(\tfrac{1}{2}, \tfrac{1}{2}; 1; \frac{1-g}{2}\right) \quad \text{when } 1 \leqslant x \leqslant 2,$$

$$= \frac{3}{2(9x-17)} F\left(\tfrac{1}{4}, \tfrac{3}{4}; 1; \frac{1}{g^2}\right) \quad \text{when } 2 \leqslant x \leqslant \tfrac{7}{3},$$

where

$$g = \frac{9x-17}{(7-3x)^{\frac{3}{2}}}.$$

The moments of $w = \frac{1}{4}b_2 - \frac{1}{4}$ about the origin are given by

$$\nu_k(w) = \frac{1}{2\Gamma\left(\frac{4k+3}{2}\right)}\left\{\Gamma(k+\tfrac{1}{2})\,\Gamma(k+1) + \frac{1}{2}\frac{k\Gamma(k-\frac{1}{2})\,\Gamma(k+2)}{4(1!)^2} + \frac{1}{2}\cdot\frac{3}{2}\frac{k(k-1)\,\Gamma(k-\frac{3}{2})\,\Gamma(k+3)}{4^2(2!)^2} + \dots\right\}.$$

Hence

$$\nu_1(w) = \frac{1}{5},$$

$$\nu_2(w) = \frac{1}{3.7} = \frac{1}{21},$$

$$\nu_3(w) = \frac{61}{5.7.11.13},$$

$$\nu_4(w) = \frac{277}{5.7.11.13.17},$$

$$\nu_5(w) = \frac{79}{3.7.13.17.19},$$

$$\nu_6(w) = \frac{46{,}889}{5.7.11.13.17.19.23}.$$

The moments about the mean of w are given by

$$\mu_2(w) = \frac{4}{3.5^2.7} = \frac{2^2}{3.5^2.7},$$

$$\mu_3(w) = \frac{-48}{5^2.7.11.13} = \frac{-3.2^4}{5^2.7.11.13},$$

$$\mu_4(w) = \frac{1424}{5^4.7.11.13.17} = \frac{2^4.89}{5^4.7.11.13.17},$$

$$\mu_5(w) = \frac{-42{,}624}{5^5.3.17.19} = \frac{-2^7.3.37}{5^5.17.19},$$

$$\mu_6(w) = \frac{76{,}096}{5^5.7.11.13.17.19.23} = \frac{2^6.1189}{5^5.7.11.13.17.19.23}.$$

Thus for the moments of b_2, we have

$$\mu_2(b_2) = 4^2\mu_2(w) = \frac{2^6}{3.5^2.7},$$

$$\mu_3(b_2) = 4^3\mu_3(w) = \frac{-2^{10}.3}{5^3.7.11.13},$$

$$\mu_4(b_2) = 4^4\mu_4(w) = \frac{2^{12}.89}{5^4.7.11.13.17},$$

$$\mu_5(b_2) = 4^5\mu_5(w) = \frac{-2^{17}.3.37}{5^5.7.11.13.17.19},$$

$$\mu_6(b_2) = 4^6\mu_6(w) = \frac{2^{18}.1189}{5^5.7.11.13.17.19.23}.$$

It may be easily verified that the same results are obtained by putting $n = 4$ in the formulae given at the end of the preceding section. There is thus reason to believe that these formulae are in fact correct.

REFERENCES

FISHER, R. A. (1929). *Proc. Lond. Math. Soc.* **30**, 199.

—— (1930). *Proc. Roy. Soc.* A, **130**, 16.

FISHER, R. A. & WISHART, J. (1931). *Proc. Lond. Math. Soc.* (2), **33**, 195.

MCKAY, A. T. (1933). *Biometrika*, **25**, 411.

PEARSON, E. S. (1930). *Biometrika*, **22**, 239.

WISHART, J. (1930). *Biometrika*, **22**, 234.

TESTING THE HOMOGENEITY OF A SET OF VARIANCES

By H. O. HARTLEY

1. Introduction

When analysing data the experimenter is frequently faced with the necessity of testing the homogeneity in a set of estimated variances. When it is desired to combine a number of variances to obtain an estimate of the common variance it is necessary to apply such a test. Again, if a selected "treatment mean square" is to be compared with an "error mean square", a test for homogeneity has recently been proposed (Wishart, 1938) as a safeguard against the selection of the largest mean square from a set of random ones.

For general use in such cases Neyman & Pearson (1931) have suggested a test: the L_1 test. The statistic L_1 used in this test has been modified by Bartlett (1937) and generalized by Welch (1935, 1936). From recent work (Nair, 1938; Bishop & Nair, 1939; Pitman, 1939) it would appear that Bartlett's statistic μ is the best to use, because it is unbiased in the sense defined by Neyman & Pearson (1936, 1938). Or more precisely, the L_1 test in its original form is biased with regard to the admissible set of alternatives.

Some difficulty has been experienced in obtaining the random sampling distribution of this statistic which is required for a test. Various approximations have been worked out. There are

(*a*) Bartlett's (1937) approximation using the χ^2 distribution.

(*b*) P. P. N. Nayer's (1936) approximation obtained by fitting Pearson-type curves to the distribution in the special case where all mean squares are based on the same number of degrees of freedom.

(*c*) U. S. Nair's (1938) expansion of the exact distribution in the special case mentioned in (*b*).

(*d*) Recently another paper on the subject has appeared by E. J. G. Pitman (1939). In this paper the author transforms the distribution of L_1 into a multiple integral which can be evaluated in special cases (small values of k) by reduction to elliptic integrals.

The accuracy of the approximations (*a*) and (*b*) has recently been tested (Bishop & Nair, 1939) in the special case in which the expansion (*c*) is available. Bartlett's findings were confirmed; it was shown that his approximation is valid only for moderate or large numbers of degrees of freedom ($\geqslant 3$). We shall also show in this paper that even with this restriction for the degrees of freedom the approximation is not very accurate if k, the number of mean squares in the set, is large.

While U. S. Nair's expansion, although it is very complicated, provides a means of working out the exact probability integral in the special case where all mean squares are based on the same number of degrees of freedom, there is still uncertainty in the general case. P. P. N. Nayer has suggested that the test for homogeneity between k mean squares with f_t degrees of freedom ($t = 1, 2, \ldots, k$) is (under certain conditions) identical with testing the homogeneity between k mean squares *all of which have f degrees of freedom*, where f is the arithmetic mean of the f_t. We shall show that, although there is some truth in this statement, the harmonic mean should be used for f rather than the arithmetic mean.

Since Bartlett's approximation does not provide a test of sufficient accuracy in all cases, the main difficulty of dealing with the general case has been the large number of quantities on which the exact distribution depends: if the k mean squares in the set have f_t degrees of freedom respectively ($t = 1, 2, \ldots, k$) the distribution would depend on $k+1$ quantities. We shall now show in this paper that (provided $f_t \geqslant 2$) there is an approximation of sufficient accuracy which depends on *three* quantities only. These three are:

(i) k, the number of mean squares in the set;

(ii) $c_1 = \sum_{t=1}^{k} \frac{1}{f_t} - \frac{1}{F}$, where $F = \sum_{t=1}^{k} f_t$;

(iii) $c_3 = \sum_{t=1}^{k} \frac{1}{f_t^3} - \frac{1}{F^3}$.

This makes the distribution amenable to tabulation, so that the test can be reduced to an inspection of a table of 5 % and 1 % points which can easily be carried out by the experimenter.

In the case where mean squares having one degree of freedom occur in the set, the distribution is of a more complicated character, but our approximation is still fair.

2. The formal solution

Consider k normal populations with variances σ_t^2 ($t = 1, 2, \ldots, k$). Let s_t^2 be an unbiased estimate of σ_t^2 based on f_t degrees of freedom, and let us denote by F the total number of degrees of freedom,

$$F = \sum_{t=1}^{k} f_t. \qquad \ldots\ldots(1)$$

Bartlett's statistic μ is then given by

$$-2\log\mu = F\log\{\sum_t (f_t s_t^2)/F\} - \sum_t f_t \log s_t^2. \qquad \ldots\ldots(2)$$

The equivalence to a special case of the generalized L_1 statistic (Welch, 1935, 1936) is expressed by the relation

$$F \log L_1' = 2\log\mu, \qquad \ldots\ldots(3)$$

where $$L_1' = \prod_{t=1}^{k} \left(\frac{F}{f_t}\right)^{f_t/F} \prod_{t=1}^{k} \left\{\frac{f_t s_t^2}{\sum_t f_t s_t^2}\right\}^{f_t/F}. \qquad \text{......(4)}$$

For our test we require the random sampling distribution $\phi(L_1')$ of the statistic L_1' under the null hypothesis

$$\sigma_t^2 = \sigma^2, \quad t = 1, 2, \ldots, k.$$

Under these conditions it has recently been shown (Welch, 1936) that the $(q-1)$th sampling moment of L_1' is given by

$$M_{q-1} = \int_0^1 \phi(L_1')\, L_1'^{q-1} dL_1'$$
$$= \prod_{t=1}^{k} \left(\frac{F}{f_t}\right)^{\{(q-1)f_t\}/F} \frac{\Gamma(\tfrac{1}{2}F)}{\Gamma(\tfrac{1}{2}F+q-1)} \prod_{t=1}^{k} \left\{\frac{\Gamma(\tfrac{1}{2}f_t + \{(q-1)f_t\}/F)}{\Gamma(\tfrac{1}{2}f_t)}\right\}. \qquad \text{......(5)}$$

From general principles it may now be inferred that equation (5) is valid for all complex q with

Real Part of $q > 1$.

Further, by Mellin's inversion formula, we obtain from (5)

$$\phi(L_1') = \Gamma(\tfrac{1}{2}F) \prod_{t=1}^{k} \Gamma(\tfrac{1}{2}f_t)^{-1}$$
$$\times \frac{1}{2\pi i}\int_{Q-i\infty}^{Q+i\infty} \prod_{t=1}^{k} \left[\left(\frac{F}{f_t}\right)^{\{f_t(q-1)\}/F} \Gamma\left[f_t\left(\frac{1}{2}+\frac{q-1}{F}\right)\right]\right] \frac{L_1'^{-q}\,dq}{\Gamma(\tfrac{1}{2}F+q-1)}, \qquad \text{......(6)}$$

where Q (>1) is an arbitrary positive quantity.

Introducing as a new statistic

$$x = -F \log L_1' = -2\log \mu, \qquad \text{......(7)}$$

and as a new variable of integration

$$\lambda = \tfrac{1}{2} + (q-1)/F,$$

we obtain for the distribution function of x ($\psi(x)$ say)

$$\psi(x) = \Gamma(\tfrac{1}{2}F) \prod_{t=1}^{k} \left(\frac{F}{f_t}\right)^{-\frac{1}{2}f_t} \Gamma(\tfrac{1}{2}f_t)^{-1} e^{-\frac{1}{2}x}$$
$$\times \frac{1}{2\pi i}\int_{A-i\infty}^{A+i\infty} \prod_{t=1}^{k} \left\{\left(\frac{F}{f_t}\right)^{\lambda f_t} \Gamma(\lambda f_t)\right\} \frac{e^{x\lambda}}{\Gamma(F\lambda)} d\lambda, \qquad \text{......(9)}$$

where A is an arbitrary positive quantity.

Using now Binet's integral representation of $\log \Gamma$ (Whittaker & Watson, 1927, p. 249), we may write equation (9) in the form

$$\psi(x) = (\tfrac{1}{2})^{\frac{1}{2}(k-1)} e^{-E(\frac{1}{2})} e^{-\frac{1}{2}x} \times \frac{1}{2\pi i}\int_{A-i\infty}^{A+i\infty} \lambda^{-\frac{1}{2}(k-1)} e^{x\lambda} e^{E(\lambda)} d\lambda, \qquad \text{......(10)}$$

where $$E(\lambda) = \int_0^\infty \left(\frac{1}{2} - \frac{1}{\tau} + \frac{1}{e^\tau - 1}\right)\frac{1}{\tau}\left[\sum_{t=1}^{k} e^{-\tau f_t \lambda} - e^{-\tau F\lambda}\right] d\tau. \qquad \text{......(11)}$$

Introducing
$$g(\tau) = \left(\frac{1}{2} - \frac{1}{\tau} + \frac{1}{e^{\tau}-1}\right)\frac{1}{\tau}, \qquad \ldots\ldots(12)$$

we see that $g(\tau)$ has continuous derivatives of any order for $0 \leqslant \tau \leqslant \infty$. We may therefore transform the integral (11) by integration by parts, differentiating $g(\tau)$ and integrating the exponential functions. We obtain

$$E(\lambda) = \frac{1}{12\lambda}\left[\sum_{t=1}^{k}\left(\frac{1}{f_t}\right) - \frac{1}{F}\right] - \frac{1}{360\lambda^3}\left[\sum_{t=1}^{k}\left(\frac{1}{f_t^3}\right) - \frac{1}{F^3}\right]$$
$$+ \frac{1}{\lambda^3}\int_0^{\infty} \frac{d^3 g(\tau)}{d\tau^3}\left[\sum_{t=1}^{k}\left(\frac{e^{-\tau\lambda f_t}}{f_t^3}\right) - \frac{e^{-\tau\lambda F}}{F^3}\right] d\tau. \qquad \ldots\ldots(13)$$

We now approximate to $E(\lambda)$, and therefore to $\psi(x)$, by ignoring the last summand in equation (13), and write

$$E(\lambda) \cong \frac{c_1}{12\lambda} - \frac{c_3}{360\lambda^3}, \qquad \ldots\ldots(14)$$

where
$$c_1 = \sum_{t=1}^{k}\left(\frac{1}{f_t}\right) - \frac{1}{F} \quad \text{and} \quad c_3 = \sum_{t=1}^{k}\left(\frac{1}{f_t^3}\right) - \frac{1}{F^3}. \qquad \ldots\ldots(15)$$

It can be shown that this approximation is sufficient for all practical purposes provided

$$f_t \geqslant 2, \quad t = 1, 2, \ldots, k.$$

Substituting (14) in (10), expanding $e^{E(\lambda)}$ and integrating the single terms we obtain*

$$\psi(x) \cong 2^{-\frac{1}{2}(k-1)}\, e^{-\frac{1}{6}c_1 + \frac{1}{45}c_3} \times \sum_{i=0}^{\infty} \alpha_i 2^{-i}\, \Gamma\left(\frac{k-1}{2} + i\right)^{-1} x^{\frac{1}{2}(k-3)+i}\, e^{-\frac{1}{2}x}, \qquad \ldots\ldots(16)$$

where the α_i are the coefficients of the expansion

$$e^{\frac{1}{6}c_1 t - \frac{1}{45}c_3 t^3} \qquad \ldots\ldots(17)$$

in ascending powers of t.

From (16) it is obvious that (to the degree of accuracy considered) the distribution of x is a weighted sum of χ^2 distributions with degrees of freedom ranging between $k-1$ and ∞. We now denote by $P_j(X)$ the probability integral of χ^2 based on j degrees of freedom, i.e. we introduce

$$P_j(X) = \Gamma\left(\frac{j}{2}\right)^{-1} 2^{-\frac{1}{2}j} \int_X^{\infty} x^{\frac{1}{2}(j-2)}\, e^{-\frac{1}{2}x}\, dx. \qquad \ldots\ldots(18)$$

We further denote by $P(X)$ the probability integral of our variate x defined in (7), viz.

$$P(X) = \int_X^{\infty} \psi(x)\, dx. \qquad \ldots\ldots(19)$$

* We make use of the well-known integral representation of $1/\Gamma(z)$, viz.

$$\{\Gamma(z)\}^{-1} = \frac{1}{2\pi i}\int_{A-i\infty}^{A+i\infty} e^{\rho}\rho^{-z}\, d\rho.$$

From equation (16) we obtain by integration

$$P(X) = \sum_{i=0}^{\infty} \alpha_i P_{k-1+2i}(X) \times \left(\sum_{i=0}^{\infty} \alpha_i \right)^{-1}, \qquad \ldots\ldots(20)$$

where the α_i are the coefficients of the expansion of

$$e^{\frac{1}{6}c_1 t - \frac{1}{45}c_3 t^3} \qquad \ldots\ldots(21)$$

in ascending powers of t.

3. Tabulation of percentage points

Equation (20) provides a means of calculating tables of the probability integral $P(X)$ (or its 5 % and 1 % points). For the quantities $P_j(X)$ are given by Elderton's tables of the probability integral of χ^2, while the coefficients α_i are readily obtained from the expansion of (21).

For practical purposes tables of the 5 % and 1 % points could be prepared. These percentage points would depend on three quantities, viz.

$$k, \quad c_1 = \sum_{t=1}^{k} \left(\frac{1}{f_t} \right) - \frac{1}{F} \quad \text{and} \quad c_3 = \sum_{t=1}^{k} \left(\frac{1}{f_t^3} \right) - \frac{1}{F^3}.$$

The effect of c_3 is small, and it would be convenient to make k and c_1 the respective row and column headings of two-way tables of percentage points, and to prepare such tables for two or three selected values of c_3. It is hoped to prepare such tables shortly.

4. Comparison with U. S. Nair's expansion

It would lead us too far afield if we gave here a complete mathematical proof of the accuracy of the approximation (20). It is, however, of interest to check the accuracy in a few cases numerically. U. S. Nair's expansion mentioned above will be used for this check. The most stringent test of the accuracy of equation (20) is established by choosing the f_t small and k large. Numerical results have been obtained from U. S. Nair's expansion (2) in the case $f_t = f = 2$; $k = 10$. The result of the test is given below.

Lower percentage points of $L_1' = e^{-X/F}$

	5% point	1% point
(*a*) Bartlett's approximation	0·367	0·277
(*c*) U. S. Nair's expansion	0·375	0·288
(*d*) Equation (20)	0·378	0·291

The agreement between U. S. Nair's expansion (*c*) and equation (20), (*d*) is satisfactory in this case where the approximation would be expected to be worst. For comparison, Bartlett's approximation (*a*) is also shown.

5. The relation between the special case $f_t = f$, $t = 1, 2, \ldots, k$ and the general case

P. P. N. Nayer has considered the general case, and has provided some evidence for believing that this case can be reduced to the special case provided that the f_t are not too small and not too dissimilar in value. He has suggested using the mean of the f_t as a substitute for the common value f. It is easy to see from the approximation (20) that there is some truth in Nayer's conjecture. However, it is not correct to use the arithmetic mean. The correct value is given by

$$f = \left(k - \frac{1}{k}\right)\left[\sum_{t=1}^{k}\left(\frac{1}{f_t}\right) - \frac{1}{F}\right]^{-1}, \qquad \ldots\ldots(22)$$

and is approximately equal to the harmonic mean of the f_t. If all $f_t \geqslant 4$, the general case of unequal f_t can *always* be reduced to the special case $f_t = f$, no matter how dissimilar the f_t. For if in equation (21) we replace c_3 by $\frac{1}{16}c_1$, and consider the function

$$e^{\frac{1}{6}c_1(t-\frac{1}{30}t^3)},$$

we find that the coefficients of this function when expanded in ascending powers of t will be approximations of sufficient accuracy to the coefficients α_t in (20). The probability integral of X is therefore determined by the quantities k and c_1, so that the identity of the general case and the special case is obvious, provided f is defined by (22).

6. Some remarks on Bartlett's approximation

Bartlett (1937) has given an approximation to the distribution of

$$-2\log\mu = x.$$

He suggests as an approximate test that we enter the table of χ^2 for $k-1$ degrees of freedom with the statistic

$$3x(k-1)/c_1,$$

where c_1 is given by (15).

It can be shown that this approximation is equivalent to equation (20) provided $\frac{1}{6}c_1$ is small, so that higher order terms in the expansion (20) may be ignored. For large or moderate values of $\frac{1}{6}c_1$, however, discrepancies may occur even if all mean squares are based on moderate or large numbers of degrees of freedom. We shall confine ourselves here to demonstrating this with the help of a single example, viz. $f = 5$ and $k = 30$. While for values of k of this order U. S. Nair's expansion is very complicated, equation (20) yields results which are accurate to 3 figures. Below are given the probabilities of exceeding Bartlett's 5 % and 1 % values; they are

	5 % level	1 % level
True $P(X)$	0·047	0·0081

Thus Bartlett's approximation has an error of 6 % and 19 % respectively.

REFERENCES

BARTLETT, M. S. (1937). *Proc. Roy. Soc.* A, **160**, 268.

BISHOP, D. T. & NAIR, U. S. (1939). *J. R. Statist. Soc. Suppl.* **6**, 89.

NAIR, U. S. (1938). *Biometrika*, **30**, 274.

NAYER, P. P. N. (1936). *Statist. Res. Mem.* **1**, 38.

NEYMAN, J. & PEARSON, E. S. (1931). *Bull. int. Acad. Cracovie*, A, 460.

—— —— (1936). *Statist. Res. Mem.* **1**, 1.

—— —— (1938). *Statist. Res. Mem.* **2**, 25.

PITMAN, E. J. G. (1939). *Biometrika*, **31**, 200.

WELCH, B. L. (1935). *Biometrika*, **27**, 145.

—— (1936). *Statist. Res. Mem.* **1**, 52.

WHITTAKER, E. T. & WATSON, G. N. (1927). *A course of modern Analysis.* 4th ed. Camb. Univ. Press.

WILKS, S. S. & THOMPSON, C. M. (1937). *Biometrika*, **29**, 124.

WISHART, J. (1938). *J. Agric. Sci.* **28**, 302.

THE SIMULTANEOUS DISTRIBUTION IN SAMPLES OF MEAN AND STANDARD DEVIATION, AND OF MEAN AND VARIANCE

By L. TRUKSA
The Charles University, Prague

In this study I propose to give the application of "the conception of the probability of passage" to the solution of the rather difficult general problem mentioned above. From this single example it is possible to deduce that the introduction of "a conception of the probability of passage" into mathematical statistics would at least make the solution of a range of difficult problems considerably easier.

Let us assume that the class symbol of the statistical element is a continuous two-dimensional variable x, y, defined in the region Ω, and that the corresponding density of the probability of passage from the class x_1, y_1 into the class x, y expressed by the symbol $p_t(x_1, y_1; x, y)$, depends, not only on the variables x, y but also on the discontinuous variable t, which is the number of operations executed on the statistical element. By operation we shall mean in our case the selection of a statistical element from the fundamental universe.

Let the function $p_t(x_1, y_1; x, y)$ satisfy the following fundamental relationship:

$$\int_\Omega \int p_t(x_1, y_1; x, y)\, dx\, dy = 1. \tag{1}$$

A further relationship, which will be used, concerns the calculations of the continuous two-dimensional probability distribution $F_{t+1}(x, y)$, corresponding to the number of operations $t+1$, from the distribution $F(x_1, y_1)$ by means of $p_t(x_1, y_1; x, y)$:

$$F_{t+1}(x, y) = \int_\Omega \int F(x_1, y_1)\, p_t(x_1, y_1; x, y)\, dx_1\, dy_1. \tag{2}$$

The problem of the simultaneous distribution of the mean and standard deviation of samples in the case in which the fundamental distribution is given quite generally by a function $f(x)$, has, so far as I know, occupied the attention of only one man; this was A. T. Craig (1932) in his study: "The simultaneous distribution of mean and standard deviation in small samples." He introduces the solution only for samples of a very small number of items, $n = 2, 3, 4$.

The use of the conception of the probability of passage enables us to demonstrate successively the solution of this problem for samples with an increasing number of items. The method used gives us at the same time a solution in a very easy manner, and especially clear, if used with a graphical illustration.

I

Let us use $\bar{x}$ for the mean of a random sample of the symbols $x_1\, x_2 \ldots x_t$ taken from the continuous one-dimensional universe of density $f(x)$, where we have

$$\bar{x} = \frac{1}{t}\sum_1^t x_i, \tag{3}$$

and for the standard deviation s, which, in accordance with the definition of the standard deviation, is given by

$$s^2 = \frac{1}{t}\sum_1^t (x_i - \bar{x})^2. \tag{3·1}$$

By an extension of this sample of size t to the size $t+2$, we get a sample, the mean of which is

$$\bar{X} = \frac{1}{t+2}\sum_1^{t+2} x_i, \tag{3·2}$$

and the standard deviation S is given by

$$S^2 = \frac{1}{t+2}\sum_1^{t+2} (x_i - \bar{X})^2. \tag{3·3}$$

The elementary probability of passage $p_t(\bar{x}, s;\, \bar{X}, S)$ from the sample with $\bar{x}$ as mean and standard deviation s to the sample of mean $\bar{X}$ and standard deviation S equals the probability of the appearance of the values x_{t+1} and x_{t+2}; thus

$$f(x_{t+1})\, dx_{t+1} f(x_{t+2})\, dx_{t+2} = p_t(\bar{x}, s;\, \bar{X}, S)\, d\bar{X}\, dS. \tag{4}$$

In this expression it is necessary to substitute for the variables x_{t+1}, x_{t+2} in terms of the variables $\bar{X}$, S; the values $\bar{x}$, s in substituting being taken as constants.

From the expressions (3) and (3·2) we obtain, first of all, the following relationship:

$$x_{t+1} + x_{t+2} = \bar{X}(t+2) - \bar{x}t. \tag{5}$$

A further relationship which we obtain from equations (3·1) and (3·3) is

$$\begin{aligned}(t+2)\, S^2 &= \sum_1^t (x_i - \bar{X})^2 + (x_{t+1} - \bar{X})^2 + (x_{t+2} - \bar{X})^2 \\ &= \sum_1^t x_i^2 - 2t\bar{X}\bar{x} + (t+2)\, \bar{X}^2 + x_{t+1}^2 + x_{t+2}^2 - 2(x_{t+1} + x_{t+2})\, \bar{X}.\end{aligned}$$

If we now use the relationship

$$ts^2 = \sum_1^t x_i^2 - t\bar{x}^2,$$

we obtain another equation, which is necessary for the determination of the values x_{t+1} and x_{t+2}, i.e.

$$x_{t+1}^2 + x_{t+2}^2 = (t+2)\,(S^2 + \bar{X}^2) - t(s^2 + \bar{x}^2). \tag{5·1}$$

From equations (5) and (5·1) it then further follows that

$$\left.\begin{aligned} x_{t+1} &= \frac{t+2}{2}\bar{X} - \frac{t}{2}\bar{x} \pm \tfrac{1}{2}\sqrt{\{2[(t+2)\,S^2 - ts^2] - t(t+2)\,(\bar{X}-\bar{x})^2\}} = \frac{t+2}{2}\bar{X} - \frac{t}{2}\bar{x} \pm \tfrac{1}{2}\alpha, \\ x_{t+2} &= \frac{t+2}{2}\bar{X} - \frac{t}{2}\bar{x} \mp \tfrac{1}{2}\sqrt{\{2[(t+2)\,S^2 - ts^2] - t(t+2)\,(\bar{X}-\bar{x})^2\}} = \frac{t+2}{2}\bar{X} - \frac{t}{2}\bar{x} \mp \tfrac{1}{2}\alpha, \end{aligned}\right\} \tag{5·2}$$

where, for the sake of brevity, we introduce the symbol

$$\alpha = \sqrt{\{2[(t+2)\,S^2 - ts^2] - t(t+2)\,(\bar{X}-\bar{x})^2\}}.$$

For effecting the substitution in expression (4) it is also necessary to know the corresponding determinant of the substitution

$$\left|\frac{D(x_{t+1}, x_{t+2})}{D(\bar{X}, S)}\right| = \frac{(t+2)^2\,S}{2\sqrt{\{2[(t+2)\,S^2 - ts^2] - t(t+2)\,(\bar{X}-\bar{x})^2\}}}.$$

Referring to the two different values for each of the symbols x_{t+1} and x_{t+2} we then get the expression for the density of probability of passage $p_t(\bar{x}, s;\, \bar{X}, S)$ in the form

$$f\left(\frac{t+2}{2}\bar{X} - \frac{t}{2}\bar{x} + \tfrac{1}{2}\alpha\right) f\left(\frac{t+2}{2}\bar{X} - \frac{t}{2}\bar{x} - \tfrac{1}{2}\alpha\right) \frac{2(t+2)^2\,S}{\alpha}. \tag{6}$$

Let $F_t(\bar{x}, s)$ be the density of the simultaneous distribution of mean $\bar{x}$ and standard deviation s in the random sample of size t.

The density of the distribution of mean $\bar{X}$ and standard deviation S in the random sample of size $t+2$ is obtained by application of the fundamental relationship (2), and for its value we get the following expression:

$$\begin{aligned} F_{t+2}(\bar{X}, S) &= \int_{\Omega}\!\!\int F_t(\bar{x}, s)\, p_t(\bar{x}, s;\, \bar{X}, S)\, d\bar{x}\, ds \\ &= 2S(t+2)^2 \int_{\Omega}\!\!\int F_t(\bar{x}, s) f\left(\frac{t+2}{2}\bar{X} - \frac{t}{2}\bar{x} + \tfrac{1}{2}\alpha\right) f\left(\frac{t+2}{2}\bar{X} - \frac{t}{2}\bar{x} - \tfrac{1}{2}\alpha\right) \frac{d\bar{x}\, ds}{\alpha}. \end{aligned} \tag{7}$$

For the complete solution of this recurrence relationship, besides the limits of integration (which we shall consider later on), we need to know the initial values of the function $F_t(\bar{X}, S)$, i.e.

$$\left.\begin{aligned} F_1(\bar{X}, S) &= f(\bar{X}), \\ F_2(\bar{X}, S) &= p_0(\bar{x}, s;\, \bar{X}, S) = 4f(\bar{X}+S)\,f(\bar{X}-S). \end{aligned}\right\} \tag{7·1}$$

Let us supplement these values with the following function $F_3(\bar{X}, S)$ for which only one integration is necessary:

$$\begin{aligned} F_3(\bar{X}, S) = 18S \int f(\bar{x})\, f[\tfrac{3}{2}\bar{X} - \tfrac{1}{2}\bar{x} + \tfrac{1}{2}\sqrt{\{6S^2 - 3(\bar{X}-\bar{x})^2\}}] \\ \times f[\tfrac{3}{2}\bar{X} - \tfrac{1}{2}\bar{x} - \tfrac{1}{2}\sqrt{\{6S^2 - 3(\bar{X}-\bar{x})^2\}}] \frac{d\bar{x}}{\sqrt{\{6S^2 - 3(\bar{X}-\bar{x})^2\}}}, \end{aligned}$$

the limits of integration will be deduced later on in this study.

In order to examine the limits of integration,

(1) *Let us first of all assume that the density of the fundamental universe is expressed by the function $f(x)$, defined between the limits $\pm\infty$.*

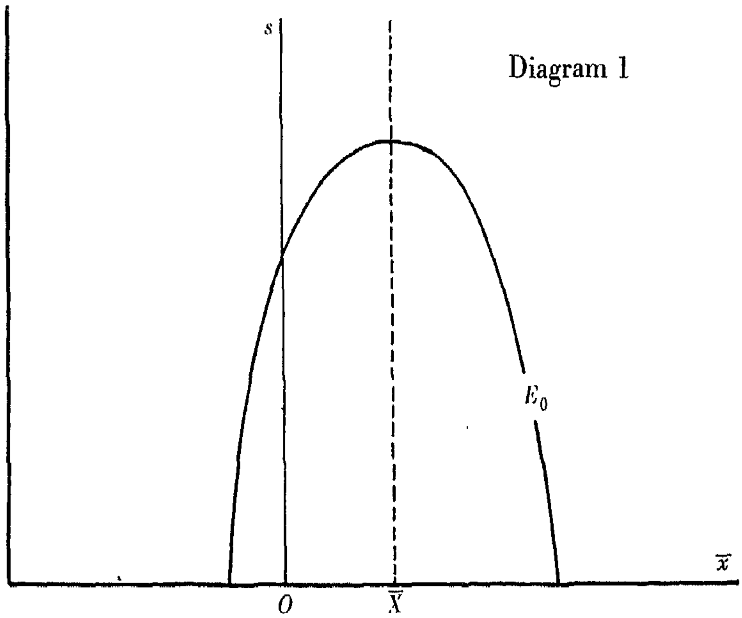
In this case we set only one condition on the values x_{t+1}, x_{t+2}, that is that they must be real. The condition corresponding to this is expressed by the inequality

$$2(t+2)\,S^2-2ts^2-t(t+2)\,(\bar{X}-\bar{x})^2\geqslant 0.$$

If we consider $\bar{X}$, S as constants and $\bar{x}$, s as variables and we use rectangular co-ordinates with axes $\bar{x}$, s, these variables are limited to the region given by that half of the ellipse E_0:

$$\frac{t+2}{2}\,(\bar{X}-\bar{x})^2+s^2=\frac{t+2}{t}\,S^2,$$

which lies above the axis $\bar{x}$. According to the choice of the values $\bar{X}$, S, the variable $\bar{x}$ varies between the limits $\pm\infty$; the variable s then lies in the range $0, \infty$.



The centre of the ellipse E_0 lies on the axis of $\bar{x}$ at a distance $\bar{X}$ from the origin of the co-ordinates; the semi-axes have the lengths

$$\lambda_s=S\sqrt{\frac{t+2}{t}},\quad \lambda_{\bar{x}}=\sqrt{\frac{2}{t}}.$$

The integration of expression (7) with respect to $\bar{x}$ must therefore be carried out between the limits

$$\left\{\bar{X}-\sqrt{\left(\frac{2S^2}{t}-\frac{2s^2}{t+2}\right)},\quad \bar{X}+\sqrt{\left(\frac{2S^2}{t}-\frac{2s^2}{t+2}\right)}\right\},$$

and then with respect to s between the limits

$$\left(0,\, S\sqrt{\frac{t+2}{t}}\right).$$

In Diagram 1 the surface of integration is shown for the particular values $\bar{X}=1$, $S=1{\cdot}5$, $t=4$.

The ellipse E_0 intersects the $\bar{x}$ axis in the points $\bar{X} \pm S\sqrt{\frac{2}{t}}$, the semi-major axis λ_s being always larger than $\lambda_{\bar{x}}$, their ratio

$$\frac{\lambda_s}{\lambda_{\bar{x}}} = \sqrt{\frac{t+2}{2}}$$

depending only on the size of the sample.

If calculation of the value of $F_3(\bar{X}, S)$, is involved, i.e. if $t = 1$, the ellipse E_0 reduces to that part of the $\bar{x}$ axis with $\bar{X} \pm S\sqrt{2}$ as end-points. The corresponding integration need only be carried out with respect to $\bar{x}$, and that between the limits

$$\bar{X} \pm S\sqrt{2}.$$

Formula (7) can be used, not only for the successive calculation of the simultaneous distribution of the mean and standard deviation for an increasing size of random samples, but also for the verification of the given distribution values introduced for the arbitrary t.

Thus, for example, it is possible to check the correctness of the expression

$$F_t(\bar{X}, S) = \frac{\sqrt{t}}{c\sqrt{(2\pi)}} e^{-\frac{t\bar{X}^2}{2c^2}} 2\left(\frac{t}{2c^2}\right)^{\frac{t-1}{2}} \frac{S^{t-2}}{\Gamma\left(\frac{t-1}{2}\right)} e^{-\frac{tS^2}{2c^2}}$$

which corresponds to the normal universe

$$f(x) = \frac{1}{c\sqrt{2\pi}} e^{-\frac{x^2}{2c^2}}.$$

By application of formula (7) we obtain

$$F_{t+2}(\bar{X}, S) = \left(\frac{t}{2}\right)^{\frac{t-1}{2}} \left(\frac{t+2}{\pi}\right)^{\frac{3}{2}} \frac{S}{c^{t+2}\Gamma\left(\frac{t-1}{2}\right)} e^{-\frac{(t+2)(\bar{X}^2+S^2)}{2c^2}}$$

$$\times \int_0^{S\sqrt{\frac{t+2}{t}}} s^{t-2}\,ds \int_{\bar{X}-\sqrt{\left(\frac{S^2}{t}-\frac{s^2}{t+2}\right)}}^{\bar{X}+\sqrt{\left(\frac{S^2}{t}-\frac{s^2}{t+2}\right)}} \frac{d\bar{x}}{\sqrt{\left(\frac{S^2}{t}-\frac{s^2}{t+2}-\frac{1}{2}(\bar{X}-\bar{x})^2\right)}}$$

$$= \left(\frac{t}{2}\right)^{\frac{t-1}{2}} \left(\frac{t+2}{\pi}\right)^{\frac{3}{2}} \frac{S}{c^{t+2}\Gamma\left(\frac{t-1}{2}\right)} e^{-\frac{(t+2)(\bar{X}^2+S^2)}{2c^2}}$$

$$\times \int_0^{S\sqrt{\frac{t+2}{t}}} s^{t-2}\,ds \left[\frac{1}{\sqrt{2}} \arcsin \frac{-\frac{1}{2}\bar{x}+\bar{X}}{\sqrt{\left(\frac{S^2}{t}-\frac{s^2}{t+2}\right)}}\right]_{\bar{X}+\sqrt{\left(\frac{S^2}{t}-\frac{s^2}{t+2}\right)}}^{\bar{X}-\sqrt{\left(\frac{S^2}{t}-\frac{s^2}{t+2}\right)}}$$

$$= \frac{\sqrt{(t+2)}}{c\sqrt{(2\pi)}} e^{-\frac{(t+2)\bar{X}^2}{2c^2}} 2\left(\frac{t+2}{2c^2}\right)^{\frac{t+1}{2}} \frac{S^t}{\Gamma\left(\frac{t+1}{2}\right)} e^{-\frac{(t+2)S^2}{2c^2}}.$$

As can be seen, the function $F_t(\bar{X}, S)$ satisfies the fundamental recurrence relationship (7).

(2) *Let the fundamental universe be defined by a continuous function defined in the range* 0, ∞.

Under this assumption the values x_{t+1}, x_{t+2}, must satisfy, apart from the condition of their being real, another condition expressed by the inequality

$$\frac{t+2}{2}\bar{X}-\frac{t}{2}\bar{x}-\tfrac{1}{2}\sqrt{\{2(t+2)S^2-2ts^2-t(t+2)(\bar{X}-\bar{x})^2\}}\geqslant 0;$$

the mean of the sample $\bar{x}$ cannot then assume negative values, which gives

$$\bar{x}\geqslant 0.$$

The corresponding values of the variables $\bar{x}$, s, satisfying the first inequality lie outside the ellipse E_1

$$\frac{t(t+1)}{2}\bar{x}^2-t(t+2)\bar{X}\bar{x}+\frac{t}{2}s^2+\frac{(t+1)(t+2)}{2}\bar{X}^2-\frac{t+2}{2}S^2=0.$$

The centre of the ellipse is situated on the $\bar{x}$ axis at a distance $\bar{X}(t+2)/(t+1)$ from the origin, the semi-axes have the lengths

$$\lambda'_s=\sqrt{\frac{t+2}{t}}\sqrt{\left(S^2-\frac{\bar{X}^2}{t+1}\right)};\quad \lambda'_{\bar{x}}=\sqrt{\frac{t+2}{t(t+1)}}\sqrt{\left(S^2-\frac{\bar{X}^2}{t+1}\right)}.$$

From the quotient $\dfrac{\lambda'_s}{\lambda'_{\bar{x}}}=\sqrt{(t+1)}$ it follows that the semi-axis λ'_s is longer than $\lambda'_{\bar{x}}$. The ellipse E_1 intersects the $\bar{x}$ axis at the points

$$\left\{\bar{X}\frac{t+2}{t+1}\pm\sqrt{\frac{t+2}{t(t+1)}}\sqrt{\left(S^2-\frac{\bar{X}^2}{t+1}\right)}\right\}.$$

The ellipse E_1 lies inside the ellipse E_0 and touches it at the point $\bar{x}_1$, s_1, where

$$\bar{x}_1=\bar{X}\frac{t+2}{t},\quad s_1=\sqrt{\frac{t+2}{t}}\sqrt{\left(S^2-\frac{2\bar{X}^2}{t}\right)}$$

as long as

$$S\geqslant\bar{X}\sqrt{\frac{2}{t}}.$$

From the expression for the lengths of the semi-axes it follows that E_1 is real only as long as the condition

$$S\geqslant\frac{\bar{X}}{\sqrt{(t+1)}}$$

is satisfied.

Considering the condition that the mean $\bar{x}$ cannot attain negative values, we obtain the inequality

$$S\leqslant\bar{X}\sqrt{(t+1)}.$$

The system of the ellipses E_1 is therefore real for those points with the parameters $\bar{X}$, S as co-ordinates, which lie between the straight lines

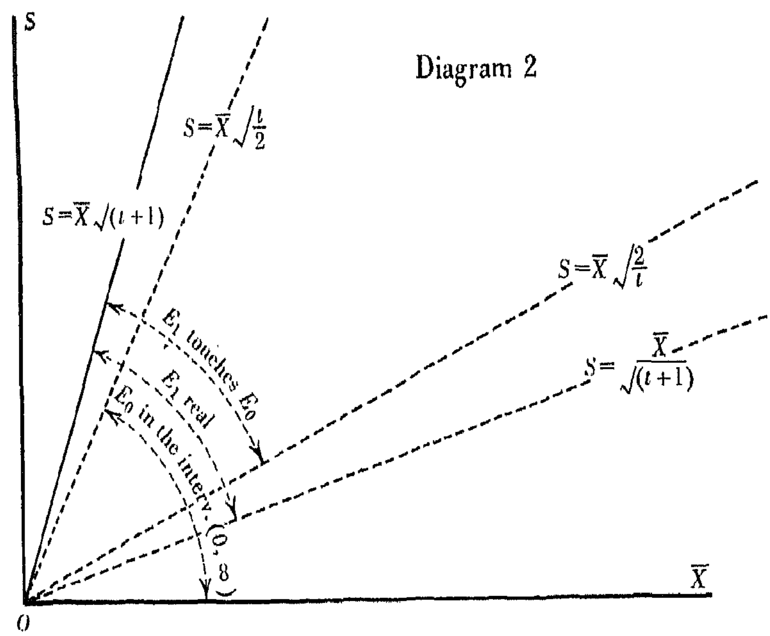
$$S=\frac{\bar{X}}{\sqrt{(t+1)}};\quad S=\bar{X}\sqrt{(t+1)}$$

and that for $\bar{X}\geqslant 0$, $S\geqslant 0$.

When carrying out the integration indicated in formula (7), it is also necessary to distinguish between the cases when the ellipse E_0 intersects the negative part of the $\bar{x}$ axis and when it does not, i.e. if

$$S \leqslant \bar{X}\sqrt{\frac{t}{2}} \quad \text{or} \quad S \geqslant \bar{X}\sqrt{\frac{t}{2}}.$$

In Diagram 2, the corresponding regions of the values $\bar{X}$, S are shown for $t = 8$.



The surface of integration is that part of the plane $\bar{x}$, s in the first quadrant contained between the ellipses E_0 and E_1.

The process of integration is then as follows:

(*a*) Let S satisfy the inequality

$$0 \leqslant S \leqslant \frac{\bar{X}}{\sqrt{(t+1)}}.$$

In this case only the ellipse E_0 is real. It is therefore necessary to carry out the integration with respect to $\bar{x}$ in between the limits $\bar{X} \pm \sqrt{\left(\frac{2S^2}{t} - \frac{2s^2}{t+2}\right)}$ and with respect to s between the limits $0, S\sqrt{\frac{t+2}{t}}$.

In order to illustrate the theory, formula (7) will be applied to find the analytical expression of the correlation surface $F_t(\bar{X}, S)$ for samples from the fundamental universe

$$f(x) = e^{-x}; \quad 0 \leqslant x \leqslant \infty$$

over the region in the plane $\bar{X}$, S bounded by the inequality mentioned above. For $t = 2$ and 3 the results are given in A. T. Craig's paper, namely

$$F_3(\bar{X}, S) = 4e^{-2\bar{X}}; \quad F_3(\bar{X}, S) = 6\sqrt{3}\,\pi S e^{-3\bar{X}}.$$

By application of our formula (7) we obtain

$$F_4(\bar{X}, S) = 64\pi S^2 e^{-4X}; \quad F_5(\bar{X}, S) = 50\sqrt{5}\,\pi^2 S^3 e^{-5X}.$$

Regarding these expressions let us assume the general solution to be

$$F_{t+2} = H_{t+2} S^t e^{-(t+2)\bar{X}}.$$

A recurrence relationship for the coefficients H_t may be easily found from the equation

$$F_{t+2}(\bar{X}, S) = 2S(t+2)^2 \int_\Omega \int H_t s^t e^{-(t+2)\bar{x}} \frac{d\bar{x}\,ds}{\alpha} = H_t 2 S^t \pi e^{-(t+2)\bar{X}} \frac{(t+2)^{\frac{t+2}{2}}}{(t-1)t^{\frac{1}{2}t}}$$
$$= H_{t+2} S^t e^{-(t+2)\bar{X}},$$

$$H_{t+2} = H_t 2\pi \frac{(t+2)^{\frac{t+2}{2}}}{(t-1)\,t^{\frac{t}{2}}}.$$

It is then not difficult to verify that for $t \geqslant 2$

$$H_t = \frac{2\pi^{\frac{t+1}{2}} t^{\frac{t}{2}}}{\Gamma\left(\frac{t-1}{2}\right)}.$$

The function $$F_t(\bar{X}, S) \frac{2\pi^{\frac{t+1}{2}} t^{\frac{t}{2}}}{\Gamma\left(\frac{t-1}{2}\right)} S^{t-2} e^{-t\bar{X}}, \quad t \geqslant 2,$$

represents a part of the whole distribution given by the integral

$$\int_{\bar{X}=0}^{\infty} \int_{S=0}^{\frac{\bar{X}}{\sqrt{(t+1)}}} F_t(\bar{X}, S)\, d\bar{X}\, dS = \frac{2\pi^{\frac{t-1}{2}} \Gamma(t)}{t^{\frac{t}{2}}(t-1)^{\frac{t+1}{2}} \Gamma\left(\frac{t-1}{2}\right)}, \quad t \geqslant 2,$$

that is, if	$t =$ 2	3	4	5
	100 %	c. 60 %	c. 30 %	c. 13 %

of the whole distribution.

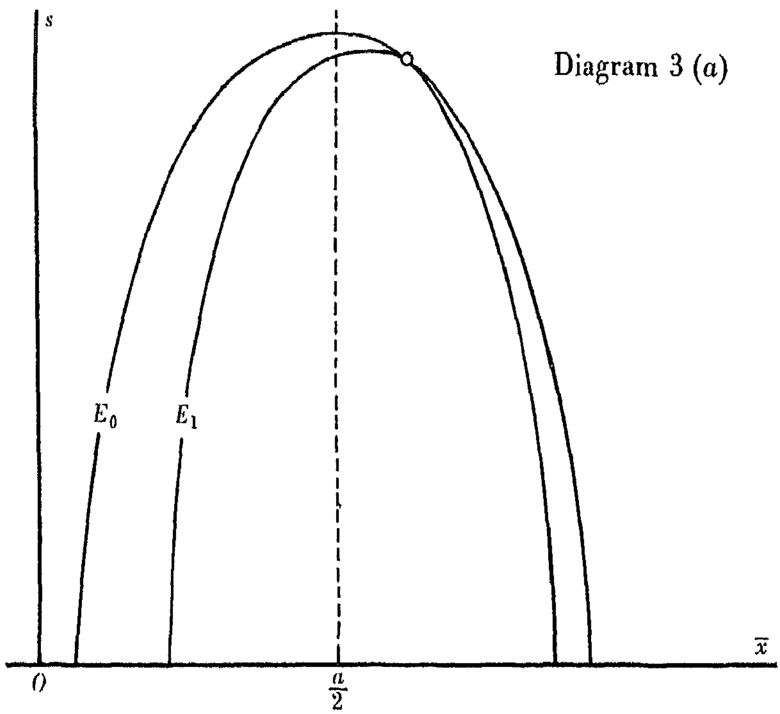
(b) If S satisfies the inequality

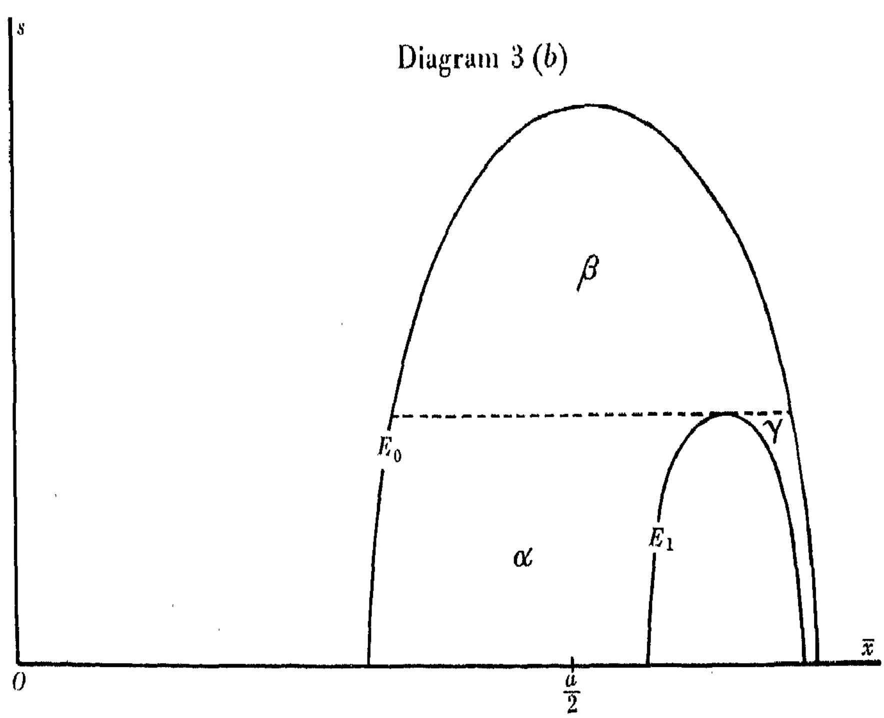
$$\frac{\bar{X}}{\sqrt{(t+1)}} \leqslant S \leqslant \bar{X} \sqrt{\left(\frac{t}{2}\right)}$$

the relative position of the ellipses E_0 and E_1 is shown in Diagrams 3a, b. The integral in formula (7) is equal to the sum of the integrals between the following limits:

(α) According to $\bar{x}$:

$$\bar{X} - \sqrt{\left(\frac{2S^2}{t} - \frac{2s^2}{t+2}\right)}; \quad \bar{X}\frac{t+2}{t+1} - \sqrt{\frac{t+2}{t(t+1)}} \sqrt{\left(S^2 - s^2\frac{t}{t+2} - \frac{\bar{X}^2}{t+1}\right)}.$$


s
Diagram 3 (a)
E_0
E_1
$\bar{x}$
0
$\frac{a}{2}$


s
Diagram 3 (b)
β
γ
E_0
α
E_1
$\bar{x}$
0
$\frac{a}{2}$

According to s:

$$0; \quad \sqrt{\frac{t+2}{t}} \sqrt{\left(S^2 - \frac{\bar{X}^2}{t+1}\right)}.$$

(β) According to $\bar{x}$:

$$\bar{X} - \sqrt{\left(\frac{2S^2}{t} - \frac{2s^2}{t+2}\right)}; \quad \bar{X} + \sqrt{\left(\frac{2S^2}{t} - \frac{2s^2}{t+2}\right)}.$$

According to s:

$$\sqrt{\frac{t+2}{t}} \sqrt{\left(S^2 \frac{\bar{X}^2}{t+1}\right)}; \quad S\sqrt{\frac{t+2}{t}}.$$

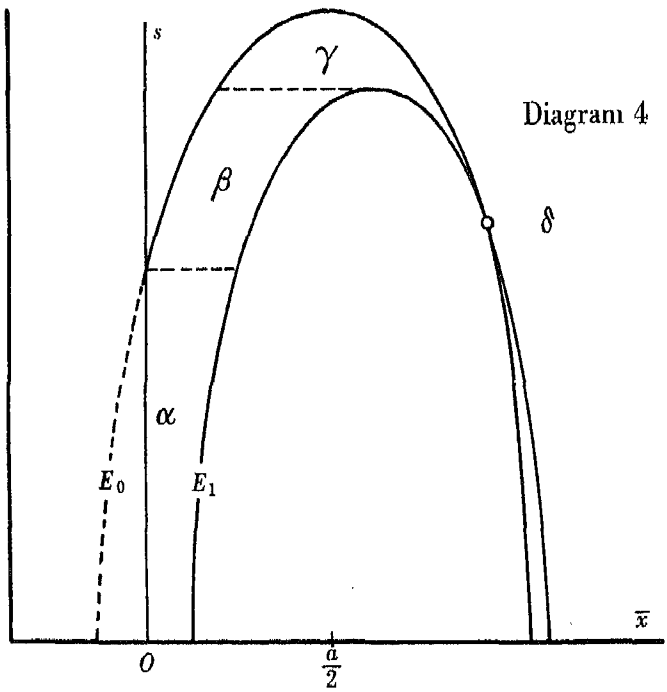


Diagram 4

(γ) According to $\bar{x}$:

$$\bar{X}\frac{t+2}{t+1} + \sqrt{\frac{t+2}{t(t+1)}} \sqrt{\left(S^2 - s^2\frac{t}{t+2} - \frac{\bar{X}^2}{t+1}\right)}; \quad \bar{X} + \sqrt{\left(\frac{2S^2}{t} - \frac{2s^2}{t+2}\right)}.$$

According to s:

$$0; \quad \sqrt{\frac{t+2}{t}} \sqrt{\left(S^2 - \frac{\bar{X}^2}{t+1}\right)}.$$

(c) Finally if $\bar{X}\sqrt{\frac{t}{2}} \leqslant S \leqslant \bar{X}\sqrt{(t+1)}$, part of the ellipse E_0 falls to the left of the s axis, E_1 always being to the right of the s axis, the surface of integration of the values $\bar{x}$, s is made up of two parts, as can be seen in Diagram 4.

For the purpose of integration it is necessary to divide the integration into four parts: $\alpha, \beta, \gamma, \delta$. The integral of the expression given in (7) is equal to the sum

of the integrals with the following limits for the value $\bar{x}$, and for s, $\bar{x}$ to be taken first of all:

(α) The limits of $\bar{x}$:

$$0;\quad \bar{X}\frac{t+2}{t+1}-\sqrt{\frac{t+2}{t(t+1)}}\sqrt{\left(S^2-\frac{s^2t}{t+2}-\frac{\bar{X}^2}{t+1}\right)}.$$

The limits of s: $\quad 0;\quad \sqrt{\left(\frac{t+2}{t}S^2-\frac{t+2}{2}\bar{X}^2\right)}.$

(β) The limits of $\bar{x}$:

$$\bar{X}-\sqrt{\left(\frac{2S^2}{t}-\frac{2s^2}{t+2}\right)};\quad \bar{X}\frac{t+2}{t+1}-\sqrt{\frac{t+2}{t(t+1)}}\sqrt{\left(S^2-\frac{s^2t}{t+2}-\frac{\bar{X}^2}{t+1}\right)}.$$

The limits of s:

$$\sqrt{\frac{t+2}{t}}\sqrt{\left(S^2-\frac{t+2}{2}\bar{X}^2\right)};\quad \sqrt{\left(S^2-\frac{\bar{X}^2}{t+1}\right)}\sqrt{\frac{t+2}{t}}.$$

(γ) The limits of $\bar{x}$:

$$\bar{X}-\sqrt{\left(\frac{2S^2}{t}-\frac{2s^2}{t+2}\right)};\quad \bar{X}+\sqrt{\left(\frac{2S^2}{t}-\frac{2s^2}{t+2}\right)}.$$

The limits of s:

$$\sqrt{\frac{t+2}{t}}\sqrt{\left(S^2-\frac{\bar{X}^2}{t+1}\right)};\quad S\sqrt{\frac{t+2}{t}}.$$

(δ) The limits of $\bar{x}$:

$$\bar{X}\frac{t+2}{t+1}+\sqrt{\frac{t+2}{t(t+1)}}\sqrt{\left(S^2-\frac{s^2t}{t+2}-\frac{\bar{X}^2}{t+1}\right)};\quad \bar{X}+\sqrt{\left(\frac{2S^2}{t}-\frac{2s^2}{t+2}\right)}.$$

The limits of s: $\quad 0;\quad \sqrt{\frac{t+2}{t}}\sqrt{\left(S^2-\frac{\bar{X}^2}{t+1}\right)}.$

If the calculation of function $F_3(\bar{X}, S)$ is especially required, that means, if $t=1$, that there is a substantial simplification in the integration. First of all it is necessary to carry out the integration only according to $\bar{x}$; besides that it is sufficient to differentiate between two cases only, according to whether the value S satisfies

$$0\leqslant S\leqslant\frac{\bar{X}}{\sqrt{2}}\quad\text{or}\quad\frac{\bar{X}}{\sqrt{2}}\leqslant S\leqslant\bar{X}\sqrt{2},$$

since obviously the case of $\frac{\bar{X}}{\sqrt{(t+1)}}\leqslant S\leqslant\bar{X}\sqrt{\frac{t}{2}}$ drops out of consideration.

If the first of the given inequalities is valid, it is necessary to carry out the integration with respect to $\bar{x}$ between the limits:

$\bar{X}-S\sqrt{2}$; $\bar{X}+S\sqrt{2}$; in the case of the validity of the second inequality, the corresponding limits of integration are:

$$0;\quad \tfrac{3}{2}\bar{X}-\tfrac{1}{2}\sqrt{(6S^2-3\bar{X}^2)},$$
$$\tfrac{3}{2}\bar{X}+\tfrac{1}{2}\sqrt{(6S^2-3\bar{X}^2)};\quad \bar{X}+S\sqrt{2}.$$

These results agree with the results quoted in the paper by A. T. Craig already cited.

(3) *Let the fundamental universe be defined in the range* 0, a.

The values x_{t+1}, x_{t+2}, in this case are determined by the inequality

$$0 \leqslant x_{t+1}, \quad x_{t+2} \leqslant a,$$

the mean of the sample satisfies the inequality

$$0 \leqslant \bar{x} \leqslant a.$$

From the upper limit of the values x_{t+1}, x_{t+2} it follows that

$$\frac{t+2}{2}\bar{X} - \frac{t}{2}\bar{x} + \tfrac{1}{2}\sqrt{\{2(t+2)\,S^2 - 2ts^2 - t(t+2)\,(\bar{X}-\bar{x})^2\}} \leqslant a.$$

The corresponding values of the variables $\bar{x}$, s, which satisfy this inequality, lie outside the ellipse E_2:

$$\frac{t(t+1)}{2}\bar{x}^2 + t\bar{x}(a - \overline{t+2}\,\bar{X}) + \frac{t}{2}s^2 = \frac{t+2}{t}S^2 - \frac{(t+1)\,(t+2)}{2}\bar{X}^2 + a(t+2)\,\bar{X} - a^2.$$

Its centre lies on the $\bar{x}$ axis and is at a distance from the origin equal to

$$\bar{X}\frac{t+2}{t+1} - \frac{a}{t+1},$$

this distance is always smaller than the distance of the centre of the ellipse E_0 from the origin.

The lengths of the semi-axes are

$$\lambda''_{\bar{x}} = \sqrt{\frac{t+2}{t(t+1)}}\sqrt{\left(S^2 - \frac{(a-\bar{X})^2}{t+1}\right)} = \frac{\lambda''_s}{\sqrt{(t+1)}}.$$

The ellipse E_2 lies inside the ellipse E_0, and touches it at the point where

$$\bar{x}_2 = \bar{X}\frac{t+2}{t} - \frac{2a}{t}; \quad s_2 = \sqrt{\frac{t+2}{t}}\sqrt{\left(S^2 - \frac{2}{t}(a-\bar{X})^2\right)}$$

as long as

$$S \geqslant (a - \bar{X})\sqrt{\frac{2}{t}}.$$

If the ellipse E_2 is to be real, then we must have

$$S \geqslant \frac{a - \bar{X}}{\sqrt{(t+1)}}.$$

From the upper limit a of the mean $\bar{x}$ of the sample it follows that

$$\bar{X}\frac{t+2}{t+1} - \frac{a}{t+1} + \sqrt{\frac{t+2}{t(t+1)}}\sqrt{\left(S^2 - \frac{(a-\bar{X}^2)}{t+1}\right)} \leqslant a.$$

This inequality leads to the relation

$$S \leqslant (a - \bar{X})\sqrt{(t+1)}.$$

Finally, it is necessary to consider the condition that the ellipses E_1 and E_2 must not intersect, in the limiting case they touch at the point

$$\bar{x} = \bar{X}\frac{t+2}{t} - \frac{a}{t}; \quad s = \sqrt{\left(\frac{t+2}{t}S^2 + 2\bar{X}\frac{t+2}{t^2}(a-\bar{X}) - a^2\frac{t+1}{t^2}\right)} = 0.$$

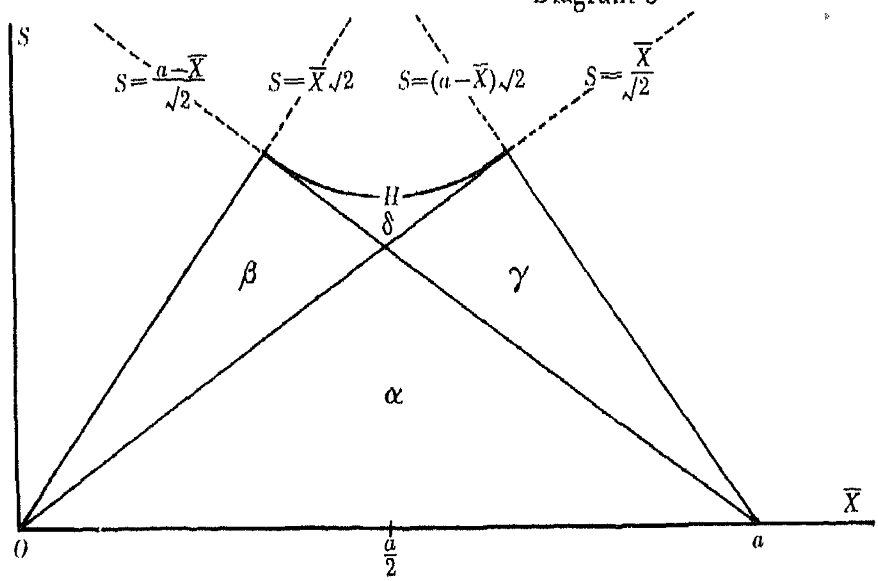
From this condition there follows the limitation of the permissible values of $\bar{X}$, S by the hyperbola H

$$S^2+2\bar{X}\frac{a\bar{X}}{t}-a^2\frac{t+1}{t(t+2)}=0.$$

For carrying out the integration in formula (7) it is necessary to know the condition for the ellipse E_0 to intersect the $\bar{x}$ axis at a distance λ from the origin, i.e.

$$S=(a-\bar{X})\sqrt{\frac{t}{2}}.$$

Diagram 5



The whole process of integration is quite clearly shown in the graphical representation of the individual segments of the permissible values $\bar{X}$, S in Diagrams 5–7, for $t=1, 2$ and 3. It is convenient to split up the surface of the values $\bar{X}$, S, for which both the ellipses E_1 and E_2 exist, into two parts by a straight line $\bar{X}=\frac{1}{2}a$.

If $$\bar{X}<\tfrac{1}{2}a,$$
the semi-axis λ'_s of the ellipse E_1 is longer than that of the ellipse E_2.

If $$\bar{X}>\tfrac{1}{2}a,$$
the ratio of the lengths of the semi-axes is the reciprocal.

The description of the limits of integration for $\bar{x}$, s is lengthy in tabular form and is better given by means of diagrams.

In the special case, $t=1$, the integration is carried out only according to $\bar{x}$, for the values $\bar{X}$, S (see Diagram 5) in the segment:

(α) Between the limits: $\bar{X}-S\sqrt{2}$; $\bar{X}+S\sqrt{2}$.

(β) Between the limits:

$$0;\ \tfrac{3}{2}\bar{X}-\tfrac{1}{2}\sqrt{(6S^2-3\bar{X}^2)}\ \text{ and }\ \tfrac{3}{2}\bar{X}+\tfrac{1}{2}\sqrt{(6S^2-3\bar{X}^2)};\ \bar{X}+S\sqrt{2}.$$

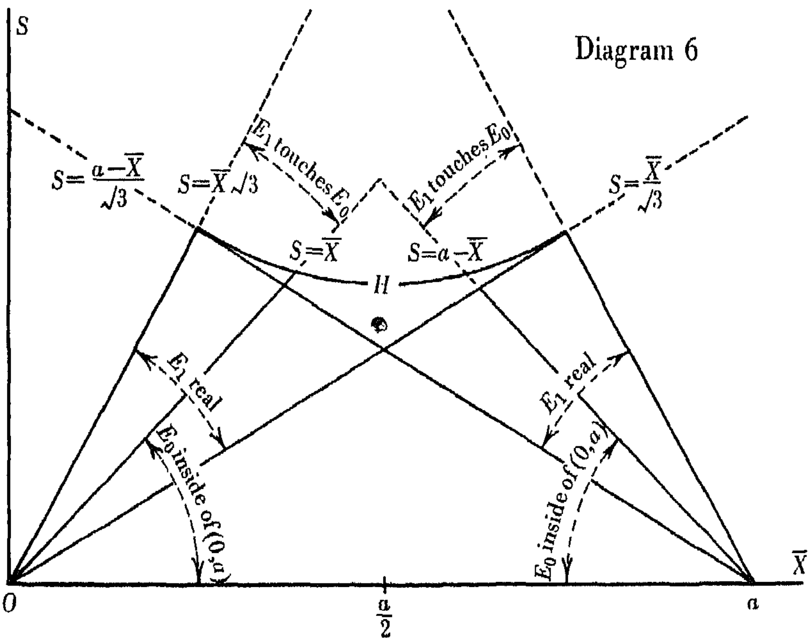
(γ) Between the limits:

$$\bar{X} - S\sqrt{2};\quad \frac{3\bar{X}-a}{2} - \tfrac{1}{2}\sqrt{\{6S^2 - 3(a-\bar{X})^2\}} \quad\text{and}\quad \frac{3\bar{X}-a}{2} + \tfrac{1}{2}\sqrt{\{6S^2 - 3(a-\bar{X})^2\}};\quad a.$$

(δ) Between the limits:

$$0;\quad \frac{3\bar{X}-a}{2} - \tfrac{1}{2}\sqrt{\{6S^2 - 3(a\bar{X})^2\}} \quad\text{and}\quad \frac{3\bar{X}-a}{2} + \tfrac{1}{2}\sqrt{\{6S^2 - 3(a\bar{X})^2\}};$$

$$\frac{3\bar{X}}{2} - \tfrac{1}{2}\sqrt{\{6S^2 - 3\bar{X}^2\}} \quad\text{and}\quad \tfrac{3}{2}\bar{X} + \tfrac{1}{2}\sqrt{\{6S^2 - 3\bar{X}^2\}};\quad a.$$



As a simple illustration let us find the correlation surface $F_t(\bar{X}, S)$ of samples of t items drawn from the distribution

$$f(x) = \frac{1}{a};\quad 0 \leqslant x \leqslant a$$

over the region bounded by the straight lines

$$S = 0,\quad S\frac{\bar{X}}{\sqrt{(t-1)}},\quad S\frac{a-\bar{X}}{\sqrt{(t-1)}}.$$

Using the results given by A. T. Craig,

$$F_2(\bar{X}, S) = \frac{4}{a^2};\quad F_3(\bar{X}, S) = \frac{6\sqrt{3}\,\pi S}{a^3},$$

we find, by our formula (7),

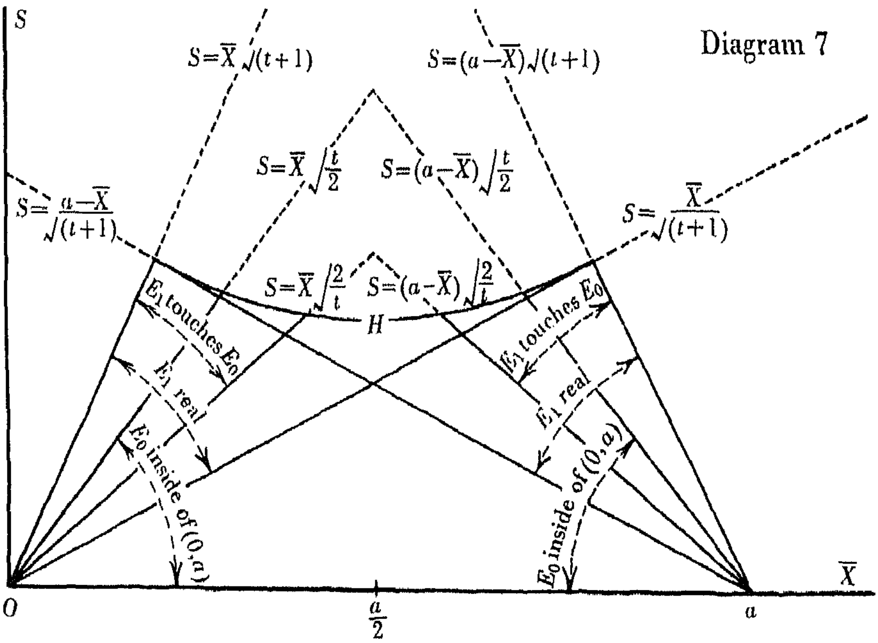
$$F_4(\bar{X}, S) = \frac{64\pi S^2}{a^4};\quad F_5(\bar{X}, S) = \frac{50\sqrt{5}\,\pi^2 S^3}{a^5},$$

and then by similar reasoning as in example on page 262 we obtain

$$F_t(\bar{X}, S) = \frac{2\pi^{\frac{t-1}{2}} t^{\frac{t}{2}}}{a^t \Gamma\left(\frac{t-1}{2}\right)} S^{t-2}.$$

This function is independent of the variable $\bar{X}$.

As to statistical *theory*, the problem of the correlation surface $F_t(\bar{X}, S)$ is solved by our formula (7) for any fundamental population $f(x)$, but as to *application* on a special distribution $f(x)$, I have not overcome all the difficulties of



integration. Nevertheless, I feel, the approach being a new one, this study may be of interest to statisticians and I hope perhaps that some mathematician will see how to solve the problems that I have left uncompleted.

II

Let us take the density of the simultaneous distribution of a mean $\bar{x}$ and variance $s^2 = u$, when the size of the sample is t, as $G_t(\bar{x}, u)$. By an application of the same process as in Part I we get the fundamental recurrence relationship for the successive calculation of the values $G_t(\bar{x}, u)$:

$$G_{t+2}(\bar{X}, U) = (t+2)^2 \int_\Omega \int G_t(\bar{x}, u) f\left(\frac{t+2}{2}\bar{X} - \frac{t}{2}\bar{x} + \tfrac{1}{2}\alpha\right) f\left(\frac{t+2}{2}\bar{X} - \frac{t}{2}\bar{x} - \tfrac{1}{2}\alpha\right) \frac{d\bar{x}\, du}{\alpha}. \tag{8}$$

The corresponding initial values are expressed by the following relationship:

$$\left.\begin{aligned} G_1(\bar{X}, U) &= f(\bar{X}) \\ G_2(\bar{X}, U) &= \frac{2}{\sqrt{U}} f(\bar{X}+\sqrt{U}) f(\bar{X}-\sqrt{U}). \end{aligned}\right\} \tag{8·1}$$

For the calculation of the next value $G_3(\bar{X}, U)$ we need to carry out only one simple integration:

$$G_3(\bar{X}, U) = 9\int f(\bar{x}) f[\tfrac{3}{2}\bar{X} - \tfrac{1}{2}\bar{x} + \tfrac{1}{2}\sqrt{\{6U - 2u - 6(\bar{X}-\bar{x})^2\}}]$$
$$\times f[\tfrac{3}{2}\bar{X} - \tfrac{1}{2}\bar{x} - \tfrac{1}{2}\sqrt{\{6U - 2u - 6(\bar{X}-\bar{x})^2\}}] \frac{d\bar{x}}{\sqrt{\{6U - 2u - 6(\bar{X}-\bar{x})^2\}}}.$$

The limits of integration for the different ranges of the fundamental universe can be deduced by the method given in Part I. At the same time the three ellipses E_0, E_1, E_2 are replaced by the three parabolae, P_0, P_1 and P_2:

$$\frac{t+2}{2}(\bar{x}-\bar{X})^2 + u = \frac{t+2}{t} U,$$

$$\frac{t(t+1)}{2}\bar{x}^2 - t(t+2)\,x\bar{X} + \frac{t}{2}u = \frac{t+2}{2}U - \frac{(t+1)(t+2)}{2}\bar{X},$$

$$\frac{t(t+1)}{2}\bar{x}^2 - t\bar{x}(a - \overline{t+2}\,\bar{X}) + \frac{t}{2}u = \frac{t+2}{2}U - \frac{(t+1)(t+2)}{2}\bar{X}^2 + a(t+2)\bar{X} - a^2$$

In conclusion I must express my thanks to Professor E. S. Pearson for advice and several useful suggestions.

REFERENCE

CRAIG, A. T. (1932). *Ann. Math. Statist.* **3**, 126.

CERTAIN PROJECTIVE DEPTH AND BREADTH MEASUREMENTS OF THE FACIAL SKELETON IN MAN

By ALETTE SCHREINER, *Oslo*

1. Definitions of the measurements

In their study of the "flatness" of the facial skeleton in man T. L. Woo & G. M. Morant (1934) lay down no method for direct measurement of the transverse flattening of the middle part of the facial skeleton, i.e. of the part made up by the malar bones and corpora of the maxillae. In studying a number of crania with differently shaped facial skeletons, it occurred to me that the best expression of the degree of projection of the middle part of the facial skeleton might be obtained by expressing the projective distances of the zygomaxillary and zygotemporal points from the most posterior points on the margins of the pyriform aperture as percentages of the lengths of the chords between the corresponding bilateral points.

It is true that no precise points of general validity for this purpose can be indicated on the "nasolateral" margins. It is also the case that we often find asymmetry here, though hardly more than in other parts of the facial skeleton on which routine measurements are taken. However, the advantages offered by the most posterior points on the margins for subtense measurements appear to me to outweigh the disadvantages. It is a factor of some importance that the two subtense planes in question are almost coincident and approximately horizontal.

After having taken some test measurements with a pair of ordinary co-ordinate callipers, I came to the conclusion that this instrument was unsuitable for my purpose. I therefore decided to undertake a preliminary research with the object of testing the value of the method, disregarding the disturbing factor of asymmetry. For this purpose I designed, and ordered from P. Hermann of Zürich, a special pair of callipers with two parallel arms, both of which might be moved in directions at right angles to the bar to give readings of the distances of the tips from the bar. According to my design the tips were to be slightly blunt, and the bar was to be only 2 mm. thick with blunt edges on its working-face. Owing to the trouble and expense involved in the construction of an instrument of this kind, P. Hermann sent for my approval one of a set of twelve pairs of callipers which he had made some years ago to the order of R. Pöch, some of which were still in his possession (Fig. 1). In most respects the design is similar to my own, but the bar is 3 mm. thick and the edges of the working-face are sharp. Furthermore, the scales of the arms are fixed in such a manner as to necessitate correction of the readings of the subtenses, whereby 4·3 mm. (controlled by the Weights and Measures Standards Office,

Oslo) must be added to the readings. In spite of these drawbacks I decided to keep the instrument, and; though it is not ideal for the purpose, it can be used effectively. I do not know for what kind of measuring it was originally designed.

For measuring a chord and its subtense I first set both arms of the instrument at equal lengths, longer than the subtense, and place the tips in contact with the extremities of the transverse chord. I then fasten the screw, draw back the arms, place the working-face of the bar lightly but firmly against the nasolateral margins, and move the arms until the tips again meet the bilateral points. On removing the instrument the distance between the arms and the subtenses are recorded, the latter after addition of 4·3 mm. If the readings on the two arms are not equal, as

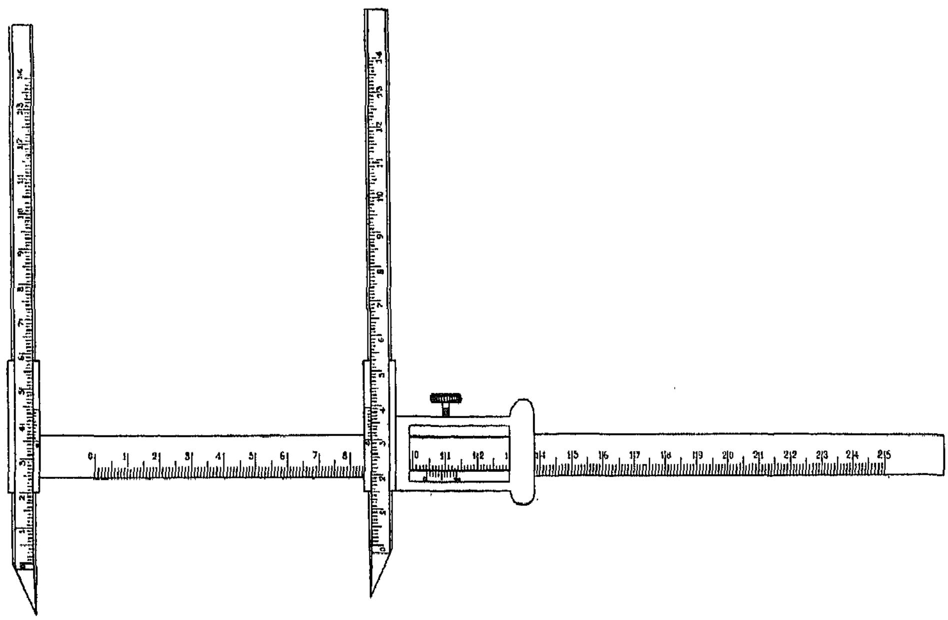


Fig. 1. Callipers used to measure the breadths and subtenses.
(The figures on the scale are centimetres.)

is most frequently the case, the average of the two values is recorded. For my method of measuring, however, I do not regard as suitable skulls with conspicuous asymmetry of the facial skeleton.

With my instrument I also took the measurements which give the "frontal index of facial flatness" of Woo & Morant in such a manner that both arms were of the same, or practically the same, length. This frequently necessitated re-measuring. I do not believe that the length of the subtense obtained by my method differs from that arrived at by using a pair of co-ordinate callipers of the usual form.

My measurements are as follows:

(i) The chord *IOW*, inner biorbital breadth, *fmo-fmo* of Martin (1928).

(ii) *Sub. IOW*, the subtense of the nasion from the chord *IOW*.

(iii) The chord *GB*, bimaxillary breadth, *zm-zm* of Martin.

(iv) *Sub. GB*, the subtense of the nasolateralia (*Nl-Nl*), i.e. the line joining the most posterior points on the margins of the pyriform aperture, from the chord *GB*.

(v) The chord *ZB*, bizygotemporal breadth, between the two *zygotemporalia inferioria* (*ZT* of Woo, 1937).

(vi) *Sub. ZB*, the subtense of *Nl-Nl* from the chord *ZB*.

(vii) *FB*, the bizygomatic breadth, *zy-zy* of Martin.

Before passing on to the indices, I must comment on the fact that the location of the zygomaxillary point (*zm*) gave me a certain amount of trouble. I did not always adhere to Martin's definition of it, which specifies absolutely the lowest point on the zygomatico-maxillary suture. Woo & Morant adopt this definition, adding that "if the inferior extremity of the suture is a short length lying parallel to the horizontal plane, the anterior point on it is the one accepted". What is the position, however, if the length is not quite short, and if it does not lie quite parallel to the horizontal plane, as when the margin is rough and irregular? The fact is that we are dealing here with a region of the facial skeleton which reveals a particularly high degree of variability, due primarily to differences in the mode of origin of the anterior part of the masseter muscle. In many cases the origin does not reach the maxilla, and where this is so location of the point presents no difficulty; but in other cases, and probably for some races in the majority of cases and particularly in males, the origin continues for different lengths on the lower border of the maxilla, and it often shows fairly strong impressions there. This last condition was frequently found in the case of the Norwegian skulls which I have examined. The same is true for the Eskimo and Australian specimens, but it was found much less frequently in the case of the skulls of Lapps and those of some other races. There appear to be racial differences in this respect.

In some cases the lowest point on the suture lies rather far back on a broad and rugged surface. Such a point would appear to be quite useless for the purpose of measuring the projection of the facial skeleton, and in fact it may not always be possible to reach it with the tip of one of the arms of the callipers when the bar is placed on the nasolateral margins. In view of this I was compelled to draw up a new definition of the zygomaxillary point, viz. the lowest point of the zygomatico-maxillary suture which still lies on the anterior surface of the bones. The adoption of this definition may occasionally alter the length of the chord *GB* by a small amount, but this is of little significance in comparison with the change which it will sometimes make in the length of the subtense.

The definition of the zygotemporal point also seems rather uncertain. Woo (1937) defines *ZT* as "the lowest point on the zygomatic suture which is still on the lateral surface of the arch". In some cases, however, it is very difficult to say where the lateral surface of the arch ends. The section of the arch shows considerable variation in form. Rectangular sections which make possible an absolutely precise location of the point are somewhat scarce. More frequently the zygoma has a lateral and a latero-inferior surface which are more or less indistinctly separ-

ated from each other. In such cases I have located the point to the best of my ability on the borderline between the two. In most cases I have found the chord to be 2–3 mm. shorter than the greatest possible breadth between the sutures on the two sides. Although the question is of much less importance here than it is for the zygomaxillary point, a more precise definition is nevertheless desirable.

Should the method which I suggest for measuring facial projections in relation to the nasolateral margins be generally adopted by craniologists, it will be necessary for them to find indisputable definitions to give the exact location of the extremities of the two chords in question.

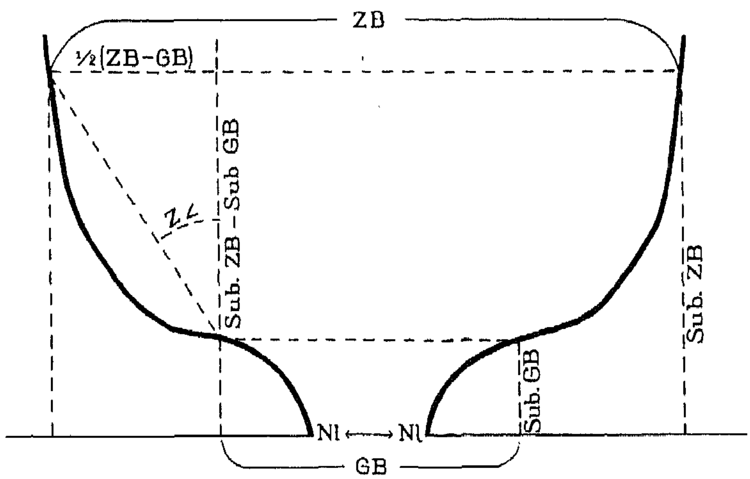


Fig. 2. A horizontal section of the facial skeleton illustrating measurements taken.

The indices used can be divided into two classes, the first (Nos. viii–xi) involving subtenses and the second (Nos. xii–xvi) being ratios of pairs of the transverse breadths. They are:

(viii) *SFi*, 100 *Sub. IOW/IOW* = frontal index of facial flatness,
(ix) *SMi*, 100 *Sub. GB/GB* = maxillary index of facial flatness,
(x) *SZi*, 100 *Sub. ZB/ZB* = zygotemporal subtense index,
(xi) *SSi*, 100 *Sub. GB/Sub. ZB*
(xii) *GOi*, 100 *GB/IOW* = maxillo-orbital breadth index,
(xiii) *ZOi*, 100 *ZB/IOW* = zygomatico-orbital breadth index,
(xiv) *GZi*, 100 *GB/ZB* = maxillo-zygotemporal breadth index,
(xv) *GFi*, 100 *GB/FB* = maxillo-facial breadth index,
(xvi) *ZFi*, 100 *ZB/FB* = zygomatico-facial breadth index.

Finally, I have calculated approximately the angle between the vertical plane through the zygomaxillary and zygotemporal points on either side and the median sagittal plane (see Fig. 2). Assuming that the two planes *zm-Nl-Nl-zm* and *ZT-Nl-Nl-ZT* are horizontal and coplanar, the fraction (*ZB-GB*)/2(*sub. ZB-sub. GB*) is the tangent of the "zygomatic angle" ($Z\angle$).

A metrical description of the curvature of the inferior margin of the malar bone would be of interest, but my attempts to measure it have been unsuccessful. A survey of such a feature would appear to be possible only by using complicated projective methods.

2. The cranial series measured

My main material consists of 100 male and 100 female Norwegian skulls from medieval churchyards in Oslo, and 100 male and 100 female skulls of Lapps obtained from various cemeteries in the county of Finmark, the majority of these being of fairly recent date. Most of the Lapp skulls form part of the material already dealt with by K. E. Schreiner (1931–5), but some of them have been acquired at a later date by the Anthropological Institute, Oslo. All of these are brachycephalic specimens. The Oslo skulls form part of the material which has been described by the same author (1939). I have looked through his records and have omitted the small number of skulls with a cephalic index greater than 79·9, or an upper facial index less than 50·0.

The series of foreign skulls in the possession of our Institute are all too small, and for the most part of too miscellaneous a nature, to provide results of any importance. I have, nevertheless, examined the best of them for the purpose of obtaining comparative data. I measured twenty-five male and twenty-five female Eskimo skulls. The majority of these came from Greenland, but two of the male and six of the female specimens are from the opposite coast of Labrador. I have found no distinct differences between the two local groups. The other series which I measured are:

Ten male and seven female Indian skulls from different parts of America, all of which have common features while none shows artificial deformation;

Eleven male and eleven female Negro skulls from different parts of Africa, all being clearly dolichocephalic;

Thirteen male and nine female native Australian skulls from different parts of the continent;

Twelve male and nine female Maori skulls from a single cave on the North Island of New Zealand.

The Australian and Maori skulls form part of the material dealt with by K. Wagner in his great work (1937). I have only included skulls sufficiently complete to provide all the measurements.

3. Sexual comparisons

Table I gives all the means which I have calculated. We will first examine the question of differences between the sexes. In Tables IIA and B sex ratios expressing the female means as percentages of the male are given for all the characters and groups, and in the lower sections of the same tables the differences between the

TABLE I

Mean measurements

Sex	No.	Series	Absolute measurements							Indices										Angle
			IOW	*Sub. IOW*	*GB*	*Sub. GB*	*ZB*	*Sub. ZB*	*FB*	*SFi*	*SMi*	*SZi*	*SSi*	*GOi*	*ZOi*	*GZi*	*GFi*	*ZFi*	*Z*∠
♂	100	Norwegian	97·71	19·23	95·61	20·21	126·98	48·27	132·84	19·68	21·14	38·01	41·79	97·85	129·95	75·30	72·02	95·60	29°·2
	100	Lapp	98·18	17·53	96·33	17·84	130·64	44·18	135·00	17·91	18·58	33·87	40·41	98·30	133·02	73·75	71·51	96·79	33°·0
	25	Eskimo	99·48	13·62	103·08	11·34	138·68	45·89	142·00	13·69	10·99	33·09	24·70	103·62	139·40	74·33	72·60	97·66	27°·5
	10	Indian	100·30	17·48	103·50	19·59	135·40	49·95	140·20	17·43	18·93	36·89	39·29	103·19	135·00	76·44	73·82	96·59	27°·7
	11	Negro	100·36	17·34	97·91	19·20	127·27	52·43	130·09	16·97	19·61	41·20	36·62	97·56	126·81	76·93	75·26	97·83	24°·4
	13	Australian	104·31	18·71	97·62	20·13	133·54	51·94	136·77	17·94	20·62	38·89	38·76	93·58	128·02	73·10	71·37	97·64	29°·6
	12	Maori	98·83	16·49	100·33	19·98	135·33	49·13	140·67	16·69	19·85	36·25	40·49	101·57	137·18	74·17	71·37	96·21	31°·1
♀	100	Norwegian	94·33	18·32	91·11	18·84	119·74	45·36	124·22	19·37	20·66	37·89	41·52	96·69	126·91	76·14	73·39	96·39	28°·2
	100	Lapp	95·14	16·29	92·24	15·89	122·81	42·67	126·51	17·12	17·23	34·74	37·24	96·91	129·08	75·11	72·91	96·92	29°·8
	25	Eskimo	95·72	13·09	98·44	10·71	128·64	44·21	131·32	13·69	10·88	34·37	24·22	102·84	134·39	76·52	74·97	97·96	24°·4
	7	Indian	94·57	15·97	97·29	18·14	124·29	46·69	129·43	16·99	18·65	37·57	38·85	102·88	131·43	78·28	75·17	96·03	25°·3
	11	Negro	94·27	15·99	91·55	19·24	116·81	47·51	119·82	16·96	21·02	40·67	40·49	97·16	123·90	78·37	76·43	98·25	24°·4
	9	Australian	97·44	16·46	88·00	16·43	122·89	48·80	125·22	16·89	18·67	39·71	33·67	90·38	126·11	71·61	70·28	98·15	28°·3
	9	Maori	95·00	14·18	95·00	17·76	127·33	46·16	131·89	14·92	18·69	36·27	38·47	100·00	133·71	74·70	72·19	96·55	29°·8

TABLE IIA

Sex ratios (female mean/male mean) and mean differences for the absolute measurements and Z∠

(i) Sex ratios

Series	*IOW*	*Sub. IOW*	*GB*	*Sub. GB*	*ZB*	*Sub. ZB*	*FB*	*Z∠*
Norwegian	96·5	95·3	95·3	93·2	94·3	93·9	93·5	96·6
Lapp	96·9	92·9	95·8	89·1	96·7	96·6	93·8	90·3
Eskimo	96·2	96·1	95·5	94·4	92·8	96·3	92·5	88·7
Indian	94·3	91·4	94·0	92·6	91·8	93·5	92·3	91·3
Negro	93·9	92·2	93·5	100·2	91·8	90·6	92·1	100·0
Australian	93·4	88·0	90·2	81·6	92·0	94·0	91·5	95·8
Maori	94·3	91·4	94·0	92·6	91·8	93·5	92·3	91·3

(ii) Sexual differences with standard errors

Series	*IOW*	*Sub. IOW*	*GB*	*Sub. GB*	*ZB*	*Sub. ZB*	*FB*	*Z∠*
Norwegian	3·38 ± 0·52	0·91 ± 0·32	4·50 ± 0·68	1·37 ± 0·40	7·24 ± 0·52	2·91 ± 0·47	8·62 ± 0·56	1°·00 ± 0·66
Lapp	3·04 ± 0·55	1·24 ± 0·30	4·09 ± 0·62	1·95 ± 0·35	7·83 ± 0·59	1·51 ± 0·46	8·49 ± 0·65	3°·2 ± 0·61
Eskimo	3·76 ± 0·90	0·53 ± 0·50	4·64 ± 0·88	0·63 ± 0·98	10·04 ± 1·39	1·68 ± 0·97	10·68 ± 1·42	3°·1 ± 0·97

TABLE IIB

Sex ratios (female mean/male mean) and mean differences for the indices

(i) Sex ratios

Series	*SFi*	*SMi*	*SZi*	*SSi*	*GOi*	*ZOi*	*GZi*	*GFi*	*ZFi*
Norwegian	98·4	97·7	99·7	99·3	98·8	97·7	101·1	101·9	100·8
Lapp	95·5	92·7	102·6	91·9	98·6	97·0	101·9	102·0	100·1
Eskimo	100·0	99·0	103·9	98·1	99·2	96·3	102·9	103·3	100·3
Indian	97·6	98·1	101·8	98·8	99·7	97·4	102·4	101·8	99·4
Negro	100·0	107·2	98·7	110·1	99·6	97·7	101·9	101·5	100·4
Australian	94·1	90·5	102·1	89·5	96·7	98·5	98·0	98·5	100·7
Maori	89·4	94·2	100·1	95·0	98·5	97·5	100·7	101·1	100·3

(ii) Sexual differences with standard errors

Series	*SFi*	*SMi*	*SZi*	*SSi*	*GOi*	*ZOi*	*GZi*	*GFi*	*ZFi*
Norwegian	$0{\cdot}31 \pm 0{\cdot}32$	$0{\cdot}48 \pm 0{\cdot}41$	$0{\cdot}12 \pm 0{\cdot}41$	$0{\cdot}27 \pm 0{\cdot}83$	$1{\cdot}16 \pm 0{\cdot}67$	$3{\cdot}04 \pm 0{\cdot}62$	$-0{\cdot}84 \pm 0{\cdot}49$	$-1{\cdot}37 \pm 0{\cdot}51$	$-0{\cdot}79 \pm 0{\cdot}23$
Lapp	$0{\cdot}79 \pm 0{\cdot}27$	$1{\cdot}35 \pm 0{\cdot}37$	$-0{\cdot}87 \pm 0{\cdot}37$	$3{\cdot}17 \pm 0{\cdot}68$	$1{\cdot}39 \pm 0{\cdot}65$	$3{\cdot}94 \pm 0{\cdot}66$	$-1{\cdot}36 \pm 0{\cdot}42$	$-1{\cdot}40 \pm 0{\cdot}46$	$-0{\cdot}13 \pm 0{\cdot}18$
Eskimo	$0{\cdot}00 \pm 0{\cdot}50$	$0{\cdot}11 \pm 0{\cdot}64$	$-1{\cdot}28 \pm 0{\cdot}74$	$0{\cdot}48 \pm 1{\cdot}34$	$0{\cdot}78 \pm 1{\cdot}28$	$5{\cdot}01 \pm 1{\cdot}04$	$-2{\cdot}19 \pm 0{\cdot}93$	$-2{\cdot}37 \pm 0{\cdot}82$	$-0{\cdot}30 \pm 0{\cdot}27$

means for the two sexes, together with the standard errors of these quantities, are provided for the Norwegian, Lapp and Eskimo series. As will be seen from Table IIA, all the male means of absolute measurements are greater than the corresponding female means, except in the case of *Sub. GB* for the Negro series. Some of the differences for the three longest series are not statistically significant. This is so for *Sub. IOW* in the case of the Norwegian and Eskimo series, although the same difference for Lapps is markedly significant. This is an unexpected conclusion and no great reliance can be placed on it, particularly in view of the fact that K. E. Schreiner has found that in general sexual differences are smaller in Lapps than in Norwegians. The peculiarity noted is undoubtedly due to the mixed composition of the Lapp material. Our Lapps do not form a homogeneous population, being mixed at different places in different degrees with Norwegians and Finns (Quains). The skulls were collected in various localities spread over a wide area. Several of the component local series are small and the sexes are unequally represented in them. I have calculated the sex ratios of this character separately for all the local groups and have found fairly large differences. Some of the group means (Kautokeino, Karasjok and Kistrand) give low sex ratios, due, presumably, to the fact that the male means are too high to be characteristic of pure Lapps.

The last column of Table IIA shows sexual comparisons of the zygomatic angle. In all series except the Negro the female mean is distinctly smaller than the male, although the difference is insignificant in the case of the Norwegian series.

Sexual comparisons of the indices (Table IIB) are of greater interest. In accordance with the results of Woo & Morant, the frontal index of flatness (*SFi*) shows no significant differences, although it appears to have a slight tendency to be lower in the female sex. This is also true for the maxillary index of flatness (*SMi*), whereas the zygotemporal subtense index (*SZi*) shows a slight tendency to be greater in the female sex (cf. the relations of $Z\angle$). The means of the *SS* index indicate, in accordance with those of the *SM* index, that the maxillary region tends to be somewhat flatter in females than in males. The values of the different breadth indices show very small sex differences, but nevertheless they show fairly consistent relationships. In all groups the biorbital breadth is a little larger relative to the bimaxillary breadth in females than in males. The three breadth measurements of the middle part of the facial skeleton (*GB*, *ZB* and *FB*) show a slight tendency to be relatively larger anteriorly than posteriorly in female compared with male skulls (cf. $Z\angle$). In my material this relation applies particularly to the Eskimos. It may be noted that for their larger series Woo & Morant found a distinctly lower mean than I have for the bimaxillary breadth (*GB*) in female Eskimo skulls. The sex ratio of this measurement in their material is only 92·7, as against 95·2 for the biorbital breadth (*IOW*).

4. RACIAL COMPARISONS

Of the breadth measurements, the internal biorbital (*IOW*) shows the smallest differences between the series and the bizygotemporal (*ZB*) the largest. If the short Negro and Australian series are disregarded. the smallest means of all the breadth measurements are found to be the Norwegian. Woo & Morant give a slightly lower mean for the biorbital breadth, and a considerably lower mean (3·4 mm. less) for the bimaxillary breadth, for their nineteen Norwegian skulls than my values for the medieval Oslo series. The values found by these authors for male Anglo-Saxon and medieval English skulls are slightly higher for the biorbital breadth, and slightly lower for the bimaxillary breadth, than my Norwegian values. With regard to the two breadth measurements in question, there is little difference between the Oslo and Lapp series, particularly in the case of the male skulls. It may be noted that for the 140 male Lapp skulls K. E. Schreiner has given a mean bimaxillary breadth of 95·5 mm., which is 0·8 mm. less than my value for male Lapp skulls and practically the same as my value for male Norwegian skulls. This author has also calculated for 121 female skulls a mean (91·7) which is slightly lower than mine. The difference between our values may be due to some extent to a difference in locating the zygomaxillary point. However that may be, the differences between my female means for *Norwegian and Lapp skulls cannot be considered statistically significant in the* case of either *IOW* ($0{\cdot}81 \pm 0{\cdot}51$) or *GB* ($1{\cdot}13 \pm 0{\cdot}56$). The bizygotemporal breadth (*ZB*), however, is significantly greater in the Lapp than in the Oslo skulls. The means of all the breadth measurements of the middle facial skeleton are clearly greater for Eskimos than for Lapps, and the biorbital breadth shows differences of the same sign, though they are much smaller. It is worthy of note that this last breadth is greater, or at all events not smaller, in the remarkably narrow-headed Eskimos than in the broad-headed Lapps. The bimaxillary breadth is probably greater in Eskimos than in any other human race, and among the known peoples of the earth only the American Indians appear to approach them in this respect.

Just as the breadth measurements increase from Norwegian to Lapp and from Lapp to Eskimo type, so, to a still greater degree, do the subtenses to *IOW* and *GB* decrease. Consequently the indices derived from these measurements show marked differences. My means of *Sub. IOW* and of the frontal index of flatness for male Oslo skulls are higher than those found by Woo & Morant for their Norwegian series, and they accord better with the values given by these authors for Swedes. On the other hand, their means of the two measurements for Eskimos are greater than mine.

With their exceedingly low means for the index of maxillary flatness (*SMi*), the Eskimo differs greatly from the other groups examined, including the Indian which otherwise bear some resemblance to the Eskimo. In fact, I know of no

characters more capable of indicating the peculiarity of the Eskimo skull than this index taken together with the chief cranial indices. Nevertheless, the Eskimos are extreme in nearly all characters, and the question remains of the extent to which the index can be considered of value for racial classification in general. I can only contribute a little towards the solution of this problem by comparing my Norwegian and Lapp skulls. The means of the index of maxillary flatness reveal the following differences: ♂ 2·56, ♀ 3·43, with standard errors of 0·40 and 0·38, respectively. The differences are of marked significance.

The zygotemporal subtense (*Sub. ZB*) and its index (*SZi*) show entirely different relations. They are associated less with the "flatness" of the facial skeleton than with the antero-posterior lengths of the calvaria, and facial skeleton. Among the groups which I have examined the Lapp skulls of both sexes have the smallest means for the subtense. The Eskimo values are distinctly higher for the subtense, but owing to its considerable zygotemporal breadth the type has the lower index. The Norwegian skulls have distinctly higher means, both for the absolute measurement and for the index, than Lapps and Eskimos, but their means are exceeded by those for the prognathous Negro and Australian skulls.

The index *SSi*, which gives *Sub. GB* as a percentage of *Sub. ZB*, shows the highest means for the Oslo skulls and, as a matter of course, very low means for the Eskimo skulls.

Among the indices which relate the different breadth measurements to one another, the maxillo-orbital breadth index (*GOi*) would appear to be of value for racial classification, as it appears to differentiate families of races. Judging from the means given by Woo & Morant for the biorbital and bimaxillary breadths the index lies below 100—that is to say, the former is greater than the latter breadth—in all European, southern Asiatic, the Australian and the majority of African populations, while it exceeds 100 in the case of eastern Asiatic, the majority of American, and a few African and Oceanic populations. As regards my material, the Lapps appear to deny their presumed Mongolian origin, since their values for this index do not exceed the Norwegian to a significant extent, while the Eskimos and Indians have means distinctly above 100.

The *ZO* index rises considerably on passing from the Norwegian to the Lapp and then to the Eskimo series. As regards the other indices of breadth, I will merely refer to the low values for the ratio of the bimaxillary to the facial breadth (*GFi*) of Lapps, who are not only distinguished by their weak mandibles, but also by their weak maxillary bones in contrast to those of Eskimos.

Finally, something must be said of the zygomatic angle. As the calculation of this is based on four different measurements, its value may be influenced by as many errors. Moreover, the two subtenses which are used do not lie in exactly the same plane. Even granting these defects, the angle, nevertheless, is fairly expressive as illustrating a feature which is not measured by the indices. I have previously dealt with the sexual differences, and will here confine myself to calling

attention to the difference between the means for Lapps and Eskimos. Both these types are characterized by broad and flat faces with what are called "high cheek bones". The low values for Eskimos may appear to be unreliable, but doubts as to this will disappear on inspection of the skull of an Eskimo from the base. The unusual breadth of the Eskimo face is due partly to the great bimaxillary breadth, and partly to the considerable curvature of the malar bones, but the posterior parts of these bones are long and the zygomatic arches only protrude slightly in a lateral direction. The bimaxillary breadth of the Lapps, on the other hand, is distinctly smaller, the malar bones are rather short in a transverse direction and the arches protrude much more laterally.

5. VARIABILITIES

In Tables IIIA and B will be found the standard deviations, while Tables IVA and B give the coefficients of variation, for the Norwegian, Lapp and Eskimo series. The bimaxillary (*GB*) is the most variable breadth and the bizygotemporal (*ZB*) is the least variable, while the bimaxillary is the most variable subtense—judging by coefficients of variation—and the bizygotemporal is the least variable. In the case of the subtense indices, also, the index of maxillary flatness (*SMi*) varies most, while the zygotemporal index varies least. These relations are undoubtedly due to the great variation in the form of the facial skeleton in the region of the zygomaxillary suture, which affects all the indices involving the bimaxillary breadth. As a test of sexual and racial differences in variability I am restricted to my Oslo and Lapp material. With regard to this question, I am bound to admit that, as long as we cannot count upon absolute accuracy in sexing skulls, very little emphasis can be laid upon any difference found. In my material male variability tends on the whole to be slightly greater than female, both as regards absolute measurements and indices, except in the case of *Sub. IOW*.

Table V gives the average coefficients of variation in the two male and two female series for three breadth measurements (*IOW*, *GB* and *ZB*), for the three corresponding subtenses, for the three corresponding subtense indices (Nos. viii–x), and for the three corresponding breadth indices (Nos. xii–xiv). Averages are also given for all six absolute measurements, for all six indices and, finally, for all twelve characters. At the bottom of the table the corresponding averages are recorded for the male plus the female constants. It will be observed that most of the male averages slightly exceed the female. As regards racial differences, the average for all six absolute measurements is greater for Lapps, but the average for all six indices is greater for Norwegians. All the differences between the averages for the two series are, however, very small.

TABLE IIIA

Standard deviations for the absolute measurements and Z∠

Sex	Series	*IOW*	*Sub. IOW*	*GB*	*Sub. GB*	*ZB*	*Sub. ZB*	*FB*	*Z∠*
♂	Norwegian	3·44 ± 0·24	2·18 ± 0·15	5·49 ± 0·39	3·15 ± 0·22	3·66 ± 0·26	3·41 ± 0·24	4·33 ± 0·31	5°·32 ± 0·38
	Lapp	4·33 ± 0·31	2·15 ± 0·15	4·77 ± 0·34	2·56 ± 0·18	4·46 ± 0·32	3·55 ± 0·25	4·80 ± 0·34	4°·52 ± 0·32
	Eskimo	3·42 ± 0·48	2·00 ± 0·28	2·52 ± 0·36	4·31 ± 0·61	5·79 ± 0·82	4·22 ± 0·60	5·90 ± 0·83	3°·84 ± 0·54
♀	Norwegian	3·96 ± 0·28	2·34 ± 0·17	3·99 ± 0·28	2·53 ± 0·18	3·70 ± 0·26	3·27 ± 0·23	3·61 ± 0·26	3°·89 ± 0·27
	Lapp	3·41 ± 0·24	2·07 ± 0·15	3·91 ± 0·28	2·43 ± 0·17	3·88 ± 0·27	3·00 ± 0·21	4·44 ± 0·32	4°·10 ± 0·29
	Eskimo	2·93 ± 0·41	1·51 ± 0·21	3·62 ± 0·51	1·83 ± 0·26	3·87 ± 0·55	2·47 ± 0·35	3·96 ± 0·56	2°·93 ± 0·41

TABLE IIIB

Standard deviations for the indices

Sex	Series	*SFi*	*SMi*	*SZi*	*SSi*	*GOi*	*ZOi*	*GZi*	*GFi*	*ZFi*
♂	Norwegian	2·21 ± 0·16	2·92 ± 0·21	3·04 ± 0·22	6·49 ± 0·46	4·94 ± 0·35	4·67 ± 0·33	4·07 ± 0·29	4·16 ± 0·29	1·96 ± 0·14
	Lapp	1·91 ± 0·14	2·77 ± 0·20	2·79 ± 0·20	4·82 ± 0·34	4·91 ± 0·35	4·77 ± 0·34	2·89 ± 0·21	3·41 ± 0·24	1·35 ± 0·10
	Eskimo	1·92 ± 0·27	2·69 ± 0·38	3·17 ± 0·45	5·36 ± 0·76	3·90 ± 0·55	4·26 ± 0·60	2·83 ± 0·40	2·47 ± 0·35	1·14 ± 0·16
♀	Norwegian	2·32 ± 0·17	2·82 ± 0·20	2·77 ± 0·20	5·22 ± 0·37	4·47 ± 0·32	4·14 ± 0·29	2·68 ± 0·19	2·87 ± 0·20	1·11 ± 0·08
	Lapp	1·91 ± 0·14	2·51 ± 0·18	2·41 ± 0·17	4·84 ± 0·34	4·23 ± 0·31	4·53 ± 0·32	3·05 ± 0·22	3·07 ± 0·22	1·20 ± 0·09
	Eskimo	1·67 ± 0·23	1·76 ± 0·23	1·95 ± 0·27	4·05 ± 0·57	5·08 ± 0·72	2·97 ± 0·42	3·08 ± 0·44	3·32 ± 0·47	0·74 ± 0·10

TABLE IVA

Coefficients of variation for the absolute measurements

Sex	Series	*IOW*	*Sub. IOW*	*GB*	*Sub. GB*	*ZB*	*Sub. ZB*	*FB*
♂	Norwegian	3·52 ± 0·25	11·34 ± 0·80	5·75 ± 0·41	15·59 ± 1·10	2·88 ± 0·20	7·07 ± 0·50	3·26 ± 0·23
	Lapp	4·41 ± 0·32	12·20 ± 0·86	4·96 ± 0·35	14·00 ± 0·99	3·41 ± 0·24	8·05 ± 0·57	3·55 ± 0·25
	Eskimo	3·44 ± 0·48	14·66 ± 2·07	2·45 ± 0·35	38·00 ± 5·39	4·19 ± 0·59	9·19 ± 1·30	4·15 ± 0·59
♀	Norwegian	4·20 ± 0·31	12·78 ± 0·90	4·39 ± 0·31	13·43 ± 0·95	3·09 ± 0·22	7·20 ± 0·51	2·91 ± 0·20
	Lapp	3·59 ± 0·25	12·68 ± 0·90	4·25 ± 0·30	15·31 ± 1·08	3·16 ± 0·22	7·04 ± 0·50	3·51 ± 0·25
	Eskimo	3·08 ± 0·44	11·52 ± 1·63	3·67 ± 0·52	17·11 ± 2·42	3·01 ± 0·42	5·29 ± 0·75	3·01 ± 0·42

TABLE IVB

Coefficients of variation for the indices

Sex	Groups	*SFi*	*SMi*	*SZi*	*SSi*	*GOi*	*ZOi*	*GZi*	*GFi*	*ZFi*
♂	Norwegian	11·24 ± 0·80	13·84 ± 0·98	7·98 ± 0·56	15·45 ± 1·09	5·04 ± 0·36	3·59 ± 0·25	5·41 ± 0·38	5·78 ± 0·41	2·06 ± 0·14
	Lapp	10·65 ± 0·75	14·93 ± 1·06	8·24 ± 0·58	11·96 ± 0·85	5·10 ± 0·36	3·59 ± 0·25	3·92 ± 0·28	4·77 ± 0·34	1·40 ± 0·10
	Eskimo	14·02 ± 1·98	24·48 ± 3·46	9·57 ± 1·35	21·68 ± 3·07	3·76 ± 0·53	3·06 ± 0·44	3·81 ± 0·54	3·40 ± 0·48	1·16 ± 0·17
♀	Norwegian	11·97 ± 0·85	13·63 ± 0·96	7·30 ± 0·52	12·56 ± 0·88	4·63 ± 0·33	3·27 ± 0·23	4·43 ± 0·31	3·11 ± 0·22	1·15 ± 0·08
	Lapp	10·69 ± 0·76	14·63 ± 1·03	6·95 ± 0·49	13·02 ± 0·92	4·36 ± 0·31	3·51 ± 0·25	4·06 ± 0·29	4·21 ± 0·30	1·26 ± 0·09
	Eskimo	12·24 ± 1·73	16·16 ± 2·28	5·66 ± 0·80	16·74 ± 2·37	4·94 ± 0·70	2·21 ± 0·31	4·02 ± 0·57	4·43 ± 0·63	0·76 ± 0·11

TABLE V

Average coefficients of variation for different sets of characters

Series	3 Breadths	3 Subtenses	3 Subt. indices	3 Br. indices	6 Absolute measurements	6 Indices	All 12 characters
Norwegian ♂	4·05	11·33	11·02	4·68	7·69	7·85	7·77
" ♀	3·39	11·14	10·97	4·11	7·52	7·54	7·53
Lapp ♂	4·92	11·42	11·27	4·20	8·17	7·74	7·95
" ♀	3·67	11·68	10·76	3·98	7·68	7·37	7·52
Norwegian + Lapp ♂	4·49	11·38	11·15	4·44	7·93	7·79	7·86
" ♀	3·78	11·41	10·87	4·05	7·60	7·45	7·53
Norwegian ♂ + ♀	3·97	11·24	11·00	4·40	7·65	7·70	7·65
Lapp ♂ + ♀	4·30	11·55	11·04	4·09	7·93	7·56	7·74

6. Conclusions

The chief result yielded by my study is that measurements of the subtenses from the most posterior points on the margins of the pyriform aperture to certain other bilateral points would be a valuable addition to the routine technique followed in describing series of skulls. Of these measurements the maxillary index of facial flatness appears to be the most useful, but the maxillo-orbital breadth index also gives comparisons of considerable interest. The adoption of these characters in craniological research will necessitate more precise definitions of the zygomaxillary and zygotemporal points than those used hitherto. Furthermore, a new instrument should be designed for measuring projections from the nasolateral margins.

References

Martin, R. (1928). *Lehrbuch der Anthropologie.*

Schreiner, K. E. (1931–5). "Zur Osteologie der Lappen." *Inst. samm. Kultur.*, Serie B, Skrifter, **18**, 1, 2.

—— (1939). "Crania Norvegica. I." *Inst. samm. Kultur.*, Serie B, Skrifter, **36**.

Wagner, K. (1937). "The craniology of the Oceanic races." *Skrift. Nor. Vid.-Akad. Oslo*, I. *Mat.-Nat. Kl.* **2**.

Woo, T. L. (1937). "A biometric study of the human malar bone." *Biometrika*, **29**, 113–23.

Woo, T. L. & Morant, G. M. (1934). "A biometric study of the 'flatness' of the facial skeleton in man." *Biometrika*, **26**, 196–250.

TRANSPOSITION OF THE VISCERA AND OTHER REVERSALS OF SYMMETRY IN MONOZYGOTIC TWINS

By E. A. COCKAYNE, D.M., F.R.C.P.

The twins, Eileen and Joan C., were first brought to my notice by Dr Reginald Lightwood, who saw them when they were six years old at the Hospital for Sick Children, Great Ormond Street, and found that one had dextrocardia and the other was normal. He remembered that they were very much alike and thought they were monozygotic. With some difficulty I got into touch with them, and Dr James Graham, Assistant County Medical Officer, Essex C.C., kindly examined the parents and the surviving brothers and sisters and found that in all of them the heart was in the normal position and no cardiac abnormality was present. The other children are: Eric, aged 21; Albert, aged 18; Doris, aged 16; Gladys, aged 12; Betty, aged 11; John, aged 9; and Iris, aged 5 years. One boy died in 1919, aged $3\frac{1}{2}$, of lymphatic leukaemia. The twins, born in August 1924, were 13 when examined. The parents are English and are not blood relations. No other case of transposition of the viscera is known to have occurred in the family.

The mother says the twins weighed 3 lb. at birth and that there were two afterbirths, i.e. each had a separate placenta. Clinically the heart is on the right in Eileen and on the left in Joan, and the size and sounds are normal in both. An electrocardiogram of Eileen taken by Dr J. L. Lovibond at the Middlesex Hospital shows inversion of all waves in lead 1, but that of Joan is normal (see pp. 290, 291). An X-ray of the chest and abdomen and a barium swallow showed dextrocardia with the stomach on the right and the liver on the left side in Eileen, and the normal position of the viscera in Joan.

The twins are very much alike in appearance and, though it is possible to distinguish one from the other when together, there is very little doubt that they are monozygotic. In view of the rarity of such a mirror image condition a number of confirmatory observations were carried out. Eileen is right-handed, while Joan is left-handed. Their handwriting is very much alike, but is unformed, and it is difficult to say how far the resemblance is due to teaching. *B*, *R*, and *r* are formed in exactly the same way in both.

Miss Ida Mann examined them and made the following report:

Eileen. Visual acuity 6/6 right and left. Hypermetropia in the right eye and hypermetropic astigmatism in the left. In both eyes the error is slightly higher than in her twin's. Retinal arterial pattern dissimilar in right and left eyes and does not resemble her twin's; practically no mesodermal pigment, well marked lesser circle and no remains of pupillary membrane. The right eye is the master eye.

Joan. Visual acuity 6/6 right and left. Low-grade hypermetropic astigmatism in both eyes. Axes oblique. Retinal arterial pattern dissimilar in right and left eyes and does not resemble her twin's. Iris pattern the same in both eyes and exactly similar to her twin's, practically no mesodermal pigment, well marked lesser circle and no remains of pupillary membrane. The left eye is the master eye.

Miss Mann says that it is usual to find the retinal arterial pattern different in the two eyes of the same person and in corresponding or opposite eyes of monozygotic twins. According to Viggo Eskelund no two persons show an identical iris pattern, but a broad classification into types is possible and these types are determined genetically. Unfortunately the iris pattern in the twins is a very common one.

Dr Phyllis Kerridge tested their hearing in a sound-proof room and both gave very similar graphs, bone conduction being normal and air conduction low normal.

Dr G. M. Morant took the following measurements in millimetres:

	Eileen		Joan	
Stature	1429		1437	
Left cubit	396		396	
Right cubit	397		396	
Maximum head length	180		184	
Maximum head breadth	136		136	
Head height (an uncertain measurement)	128		125	
Cephalic index	75·6		73·9	
Bizygomatic breadth	121		121	
	R.	L.	R.	L.
Maximum length of ear	56	56	55	57
Length of phalanges with fingers strongly flexed at metacarpophalangeal joints:				
2nd digit	92	92	94	93
3rd digit	99	101	103	102
4th digit	93	95	98	98
5th digit	76	76	75	78

The girls are about 2½ in. below the English average stature for their age and social class. The small differences between measurements suggest that they are monozygotic. The small difference between head lengths (leading to one between the cephalic indices) is probably the most significant. Joan has slightly longer digits than Eileen. As all the measurements are subject to errors of at least 1 mm. it is safest to conclude that they do not provide any evidence of asymmetry.

He also took photographs and reported as follows: The photographs are of three kinds.

(I) *Profiles*. The outlines of the faces of the two girls are remarkably similar, their upper lips being unusually short. Eileen's chin is rather longer and straighter than Joan's. No difference between the ears (not all visible) was noted. Both have free lobes. They provide no evidence of asymmetry.

(II) *Full-face*. The photographs are not truly full-face, rather more of the

right than of the left being shown in each case. The outlines are very similar, Joan having a rather shorter and more pointed chin than Eileen. The teeth show a slight anomaly, which is the same in both girls. The central upper incisors are crossed, the left overlapping the right.

A palmaris longus muscle is present in both twins on both sides. Hair colour in both twins matches scale 7. The hair whorl was clockwise in both twins.

Miss David took fingerprints and made the following report:

A glance at Waite's tables shows that radial loops are nearly as common as ulnar loops on the forefinger. The remaining fingers are all ulnar loops, which are the commonest of all patterns. There is a distinct similarity between the counts of these ulnar loops and in the actual pattern. In both twins the actual counts approximate to the median value of the distance of ridges in loops. The thumbs are interesting and at first sight appear very similar, but there is a difference as shown in the counts, and the right thumb of Eileen is like the left thumb of Joan, while the left thumb of Eileen is like the right thumb of Joan. The most striking thing is that Eileen has a radial loop on the forefinger of the right hand, and Joan has a radial loop on that of her left hand, and Eileen has an ulnar loop on the forefinger of the left hand and Joan has an ulnar loop on that of the right hand.

	Right hand		Left hand	
	Joan	Eileen	Joan	Eileen
T	23 C 9	21 C 6	5 C 20	11 C 24
1	UL 13	RL 5	RL 5	UL 13
2	UL 13	UL 14	UL 12	UL 13
3	UL 12	UL 12	UL 12	UL 13
4	UL 13	UL 12	UL 13	UL 10

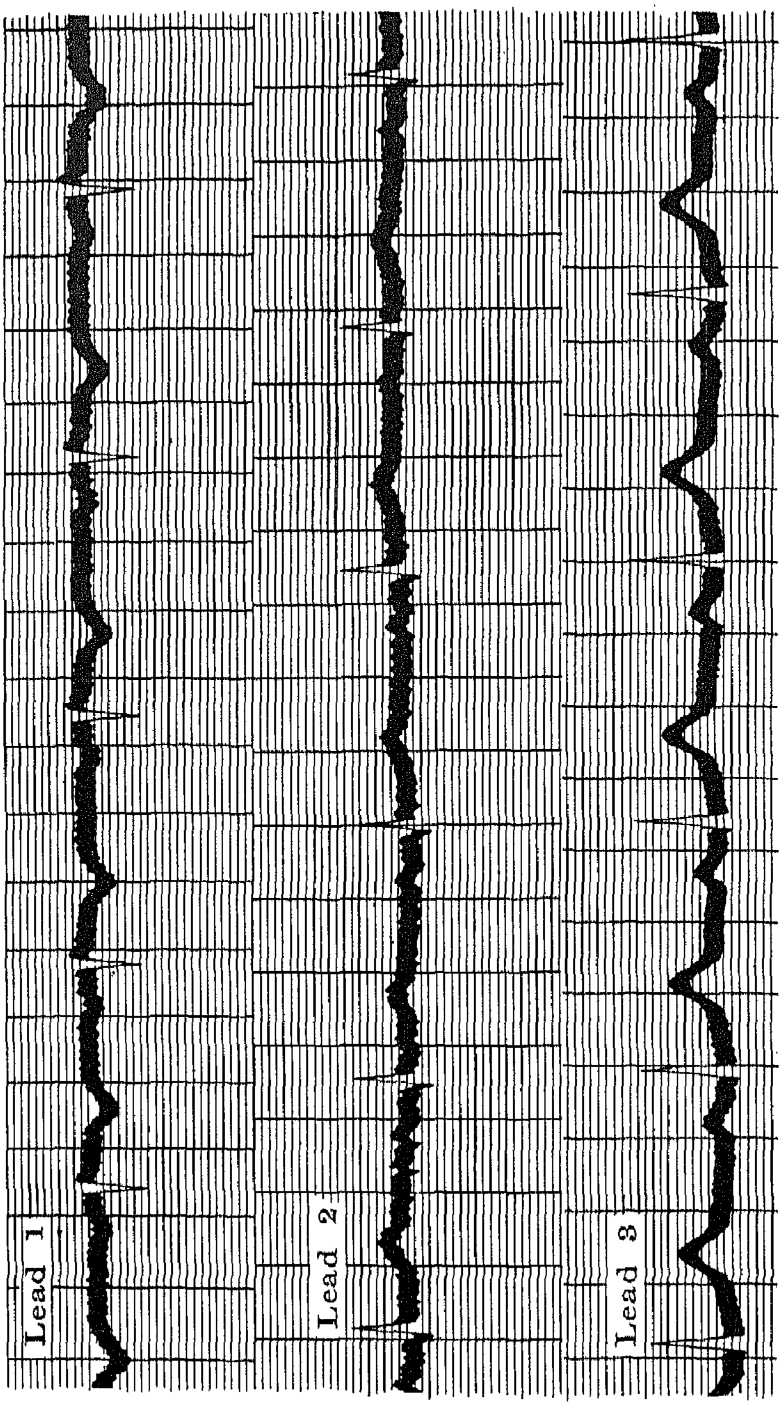
UL = Ulnar loop. RL = radial loop. C = composite.
The numbers are the number of ridges.

The palmar and plantar patterns are similar, but those of Eileen's left side are more like those of Joan's right side than those of her own right side and vice versa.

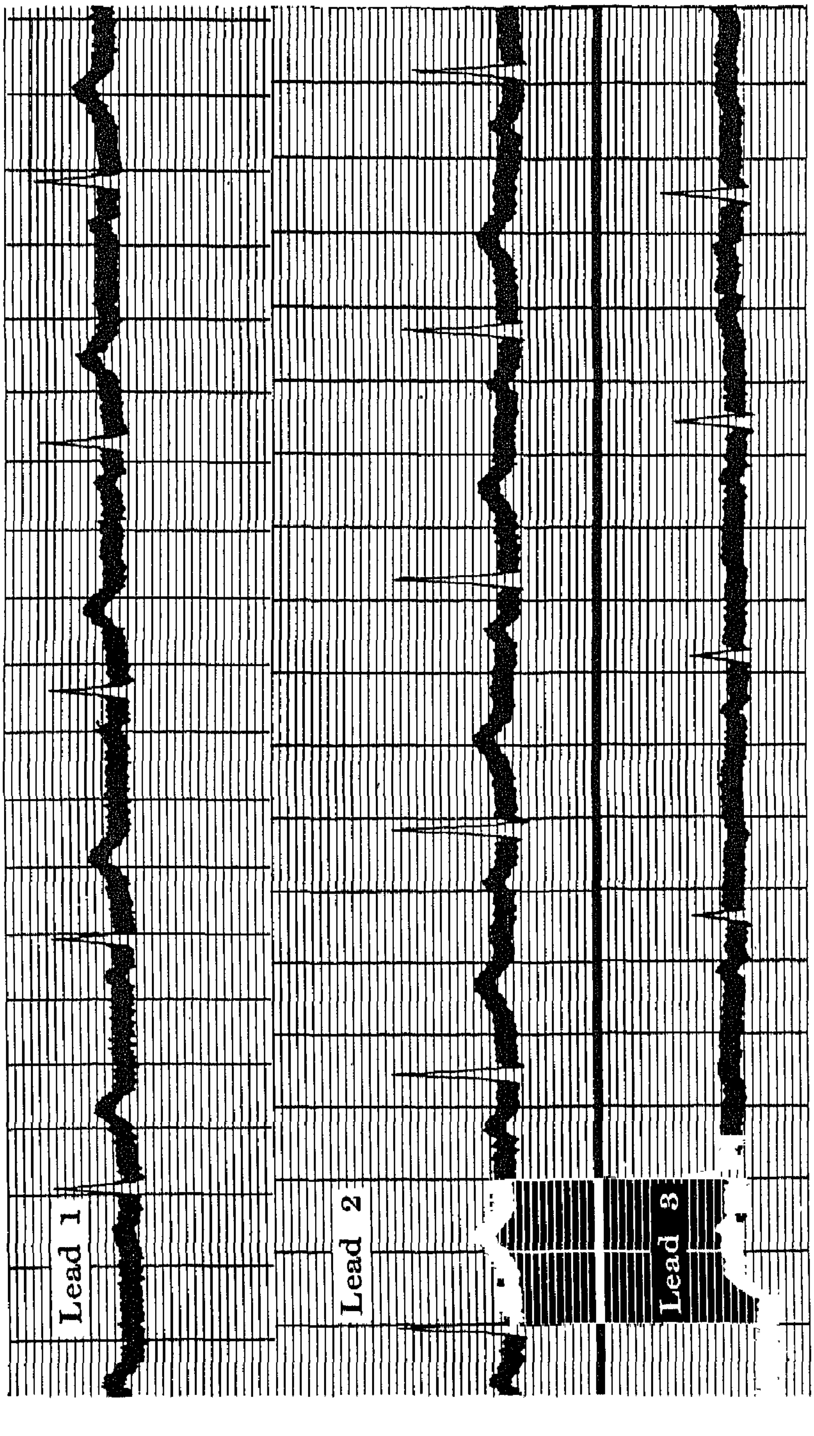
Dr G. L. Taylor finds that the blood groups in both twins are *O M N*. Both are able to taste phenyl-thio-carbamide. The other taste tests were unsatisfactory because they were unable to discriminate clearly between sour and bitter.

At school the twins were in the same class and much alike in mentality. Questioned in different rooms with no opportunity of overhearing one another or comparing notes each said that her favourite colour was blue, one immediately and the other after some hesitation. Neither knew her sister's favourite colour.

I think the resemblance in general appearance and the agreement in so many independent characters, a number of which are known to be determined genetic-



Eileen. Electrocardiograms.



Joan. Electrocardiograms.

ally, proves that Joan and Eileen are monozygotic. They differ from the majority of such twins in the extent to which they are mirror images of one another:

Eileen with complete transposition of the viscera.

Joan with normal viscera.

Eileen right-handed with the right eye the master eye.

Joan left-handed with the left eye the master eye.

Eileen with a radial loop with five ridges on the right forefinger.

Joan with an radial loop with five ridges on the left forefinger.

Eileen with an ulnar loop with thirteen ridges on the left forefinger.

Joan with an ulnar loop with thirteen ridges on the right forefinger.

Eileen with thumb ridges of the right hand closely resembling those of Joan's left hand, and with thumb ridges of the left hand closely resembling those of Joan's right hand.

I can find records of only four similar cases in the literature. Baron (1825) showed one before the Section of Medicine of the Academy of Medicine, Paris, in December 1825. A short report says that the twin with transposed viscera was a girl, who died at the age of 8 days. Kuchenmeister states that the twins were monozygotic and that one had normal viscera and the other complete transposition. Miller (1893) records a case of monozygotic twins, one with situs inversus viscerum and the other with situs solitus, but gives neither their age nor sex. Dubreuil-Chambardel (1927) records male twins, aged 25, identical in general appearance, in weight, height, and in physical characters. Each had varices on the lower limbs of the same type and developing in the same way. One had a harelip on the left, the other on the right, and one had complete transposition of the viscera, while the other was normal in this respect. The visceral condition was confirmed by radiological examination. I have been unable to refer to the original account of the twins recorded by Araki (1934), but Taku Komai, though he gives neither their age nor sex, says they were monozygotic and that one had normal viscera and the other complete transposition. Taku Komai says that this is the fifth case recorded. Possibly he includes that recorded by Tamm and later by Betschler. This refers to female twins, stillborn at the seventh month, one normal, the other oedematous and hydrocephalic, with the heart and aorta on the right side, no lung on the right and a small one on the left. The other viscera were correctly placed, but the liver was small and the kidneys large. There was one placenta with two cords implanted centrally with the umbilical veins anastomosing.

Newman in his *Biology of Twins* shows that the armadillo, *Dasypus novemcinctus*, normally produces monozygotic quadruplets. He makes the following observations. At first there are two embryos and later a secondary embryo is formed to the left of each primary embryo. Asymmetry of the scutes is not uncommon and in the quadruplets all grades of mirror image formation are found. There may be mirror-imaging between individuals of opposite pairs (primary embryos), but this is much less common than imaging between twin

partners derived from one half of the egg (a primary embryo and the secondary embryo to its left). In general mirror-imaging between "opposites" is evidence of a residuum of a primary bilateral symmetry that held sway in the blastocyst before polyembryonic fission began. When the primary outgrowths are formed, they are the product of the antimeric halves of the first embryo and should therefore show mirror image relations. But a partial physiological isolation of the two halves permits a certain reorganization or regulation of new symmetry relations, which tends more or less completely to destroy the original symmetry, yet often leaving a trace of the latter. Finally when each secondary outgrowth organizes its own bilateral symmetry, it tends to lose, partially at least, the earlier symmetry relations, and to establish its own mirror-imaging of right and left halves. In some cases traces of all three symmetry systems appear in a single set of foetuses, but it is common to find only two systems interacting. Newman says that no case of visceral reversal was found in spite of a careful search through a large number of foetuses. This, he thinks, is due to the fact that twinning is initiated and carried out in the ectoderm, and that the endoderm becomes involved only passively and considerably later.

In the case of human monozygotic twins there may be almost complete mirror image arrangement of palmar patterns so that the left hand of x corresponds with the right hand of y and vice versa. This corresponds with the rather rare mirror-imaging of "opposites" seen in armadillo quadruplets and in Newman's opinion goes far to prove that polyembryony actually occurs in man. A commoner manifestation in man is reversal of symmetry in the pattern of an index finger of one twin so that it mirrors the condition in the corresponding index finger of the other twin. This reversal of pattern occurs in both hands of Eileen and Joan.

In addition to the ridges on the skin other epidermal structures may show a mirror image arrangement, such as naevi and angiomata, and Siemens (1924) records an accessory nipple below the normal one, which was on the right side in one and on the left side in the other twin.

Newman believes that monozygotic human twins become physiologically isolated at a considerably earlier period than do armadillo quadruplets and founds his belief on the fact that there is so little mirror-imaging in the former and so much in the latter. He says that as a general rule the earlier the separation the more complete is the reorganization of the symmetry relations in the separate individuals and the less residuum of the original common symmetry.

In double monsters of monozygotic origin there is, according to Dubreuil-Chambardel (1927), frequently a hare-lip, which is axial in both or lateral in both, and as a rule it is of the same degree, and the hare-lip of one component is always a mirror image of the hare-lip of the other. He points out as many others have done how often one component of a double monster has complete or partial transposition of the viscera, the arrangement being that the apices of the two hearts point away from the place where the twins are attached to one another.

There can be no doubt that neither the hare-lip nor the transposition of viscera in double monsters is of genetic origin. Hare-lip in man is sometimes determined genetically, but the incidence in double monsters is far too high for this to be true in their case. Newman says that in the case of double monsters fission must begin much later than usual and is never completed. While in the armadillo polyembryony is the rule and fission is standardized, in man it is the exception and there is considerable variability in the period at which fission takes place. Transposition of the viscera in man is recessive to the normal arrangement, and elsewhere I have suggested (1938) two ways in which it could affect only one member of a pair of monozygotic twins. Apart from the improbability of so unusual an occurrence happening in at least five recorded cases, neither explanation accounts for the occurrence of other mirror image arrangements, and any genetic explanation must be abandoned. If Newman's views about mirror image formation are accepted, the very rare cases of human monozygotic twins with a mirror image arrangement of the viscera in addition to that of ectodermal structures are due to an unusually late fission. The fact that they are so rare seems to show that unless fission is very late each twin reorganizes its own symmetry and, if genetically normal, the viscera are normal in both, but if the zygote is a homozygous recessive for transposition of the viscera, the viscera are transposed in both. Apparently few twins in which late fission occurs succeed in becoming completely separate, for double monsters with their own mirror image symmetry, i.e. with the viscera transposed in one and normal in the other, are far commoner than separate twins with this arrangement. The separate twins, which do show this arrangement have in fact just escaped being double monsters.

My thanks are due to Prof. R. A. Fisher for kindly putting the resources of the Galton Laboratory at my service and to Dr Julia Bell for her help and advice, and to Miss David, Dr Taylor and Dr Morant. I wish also to express my gratitude to Miss Ida Mann and to Dr Lovibond. Dr Graham had intended to write a report on the twins, and I should like to thank him for allowing me to do so instead and for examining the other members of the family.

REFERENCES

ARAKI, B. (1934). *Nagasaki Igakkai Zasshi*, **13**, 11, 1691.

BARON, J.-F. (1825). *Bull. sci. Med.* (1826), **8**, 132. Cited by Küchenmeister, F., *Verlagerung der Eingeweide des Menschen*, Leipzig, 1883, p. 125.

BETSCHLER. Cited by Schatz, F. *Arch. Gynäk.* (1887), **29**, 425.

COCKAYNE, E. A. (1938). "The genetics of transposition of the viscera." *Quart. J. Med.* N.S. **7**, 479.

DUBREUIL-CHAMBARDEL, L. (1927). *Pr. Med.* **35**, 877, 1157.

MILLER, N. (1893). *Jb. Kinderheilk.* **36**, 340.

NEWMAN, H. H. (1917). *The Biology of Twins.* Chicago.

SIEMENS, H. W. (1924). *Die Zwillingspathologie*, p. 43. Berlin.

TAKU KOMAI (1937). *Studies on Japanese Twins*, p. 6. Kyoto.

Eileen.

Joan.

Eileen.

Joan.

Cockayne: *Transposition of the viscera and other reversals of symmetry in monozygotic twins*

Joan.

Eileen.

THE HUMAN REMAINS OF THE IRON AGE AND OTHER PERIODS FROM MAIDEN CASTLE, DORSET

By C. N. GOODMAN and G. M. MORANT

1. Introduction

Excavations at Maiden Castle, the largest fortified site of prehistoric date in England, were conducted under the direction of Dr R. E. M. Wheeler during four consecutive seasons from 1934 to 1937. The discoveries made were of unusual archaeological interest, and they include a number of skeletal remains which were preserved with unusual care. The bones available represent 104 individuals ranging in age from a foetal to a senile stage of development. This paper is a report on the eighty-three skeletons which are sufficiently well preserved to be of some anthropological interest, and the periods, sexes and age groups to which they are assigned are given in Table I. An unpublished report on the excavations by Dr Wheeler contains particulars of each of the 104 individuals, and a section by the present writers in the same work gives a summary of the general conclusions reached below and fuller descriptions of the mutilated specimens.

TABLE I

The periods, sexes and ages of individuals from Maiden Castle whose skeletons were measured

Period or group	Adult ♂	Adult ♀	Adolescent ♂	Adolescent ♀	Infant	Totals
Neolithic	3	1	—	—	—	4
Iron Age *A*	1	—	—	—	4	5
Iron Age *B*	2	7	1	—	3	13
Iron Age *C*	8	8*	—	1	—	17
Belgic	—	1	—	—	1	2
Belgic War Cemetery	20	9	—	1	1	31
Romano-British	3	3	—	—	—	6
Iron Age or Romano-British	3	1	—	—	—	4
Saxon	1	—	—	—	—	1
Totals	41	30	1	2	9	83

* Including one individual (T 28) classed as Iron Age *C* or Romano-British.

The metrical study of the skeletons made hitherto has been confined to the crania and mandibles, on which all the customary measurements have been taken, and to the clavicles and long bones of the arms and legs, described by

lengths only. Individual measurements and remarks for the crania and long bones are given in the appended Tables VII and VIII; remarks on the mandibles are given with those on the crania, but the individual measurements of these bones are not provided. Contours of the crania have been drawn, but they are not used here, as additional English material of the same periods would be needed to furnish sufficiently reliable type figures. The writers hope to undertake a more detailed examination of the long bones and an examination of the other bones of the skeleton.

TABLE II

Maximum lengths of the shafts of long bones of infants of Iron Age date and of modern foetuses

No. ...		GM 1	P III	Q II	R 1	P 1	R II	T 19	O	P 35	—	—
Year excavated ...		1936	1937	1937	1937	1937	1937	1937	1936	1937	—	—
Period* ...		*A*	*A*	*A*	*A*	*B*	*B*	*B*	*C*†	*C*‡	Modern§	Modern‖
Femora	R.	—	75·1	—	78·9	—	—	—	—	84·9	65·1	58·0
	L.	—	75·1	—	—	85·2	76·9	72·4	81·5	—	65·0	58·0
Tibiae	R.	—	64·0	—	68·2	—	67·0	68·2	—	—	56·1	51·0
	L.	—	64·2	—	68·8	—	67·1	68·8	—	72·1	56·5	51·1
Fibulae	R.	—	61·0	—	—	—	—	61·2	68·0	—	—	48·2
	L.	—	61·2	—	—	—	—	61·2	—	—	53·0	48·2
Humeri	R.	70·5	64·1	—	66·0	74·0	67·0	66·3	71·4	73·0	58·1	53·1
	L.	70·7	64·8	75·5	67·8	—	—	—	71·4	73·0	58·2	53·0
Radii	R.	—	49·0	—	54·4	58·2	52·7	51·8	55·5	—	45·0	44·0
	L.	—	49·1	—	54·4	58·0	—	51·2	55·7	57·9	45·0	44·0
Ulnae	R.	64·4	58·2	—	—	66·2	61·1	—	—	—	53·0	49·0
	L.	—	58·0	—	61·4	—	—	58·8	64·5	—	53·0	49·0
Clavicles	R.	—	46·0	—	—	48·3	—	42·1	49·1	46·4	39·2	34·5
	L.	47·0	46·0	—	—	48·4	—	42·6	—	—	38·8	34·0

* *A*, *B* and *C* refer to the chronological divisions of the Iron Age. † Belgic.
‡ Belgic War Cemetery. § Full term. ‖ 7 months foetus.

Nearly all the skeletons are incomplete, but in the majority of cases the sexes of the adult and adolescent individuals can be judged, principally from the pelvis, with a reasonable assurance of accuracy. It was suspected at first that nine skeletons, which are remarkably well preserved in view of their tender age, were foetal. The maximum lengths of the shafts of the long bones for these—the measurements having been taken in any direction by using the flat arms of a pair of small calipers—are given in Table II. The last two columns on the right of the table give the same lengths for a child at birth and a seven months foetus. These two specimens are preserved in the museum of the Department of Anatomy, University College, London, and we are indebted to Dr Matthew Young for giving us access to them. All nine of the skeletons from Maiden Castle are appreciably larger than the two of modern date, and hence it must be

supposed that the former are of young infants, probably all from three to six months old. No other measurements of them were taken.

The remaining seventy-four skeletons are available for racial comparisons, and the majority of these are of Iron Age date. Thirty represent the population of the site at the very end of the period, as they came from a cemetery used for the Belgic defenders of the Castle who were slain by Roman invaders in the year A.D. 43. Nearly half of the skulls in this series bear witness to the event in the form of sword-cuts and other mutilations. The Belgic War Cemetery series is just long enough to make statistical comparisons between it and other groups worth while, but all the other groups distinguished in Table I are clearly too small to stand alone. In some comparisons below a composite series made up by pooling all the other individuals of Iron Age or Romano-British date is considered, either apart from the Belgic War Cemetery series or combined with it. This procedure is of provisional value only, of course. There may have been changes in the racial composition of the population of Maiden Castle during the Iron Age and again in Roman times, and the evidence available is quite insufficient to show whether this was so or not.

It should be appreciated, too, that it is not unlikely that a small community, which may have been peculiar owing to relative inbreeding, may have persisted there for several generations, and the characteristics of such a local group may mislead if it is taken to typify the large racial population of which it formed a special part. The new material from Maiden Castle provides a very welcome addition to our meagre knowledge of the physical characters of the inhabitants of England in Iron Age times. It may be hoped that it will form a nucleus to which other specimens of the period—both those already housed in museums and others as yet undiscovered—may be added, until the evidence is abundant enough to satisfy the most exacting anthropologist. The object of this paper is to provide a descriptive record of the skeletons available with such a hope in view, and any results regarding racial relationships made in it are intended to be of a tentative nature only.

2. NON-METRICAL FEATURES OF THE MAIDEN CASTLE SERIES

The remarks on the Iron Age and Romano-British series given in this section relate almost entirely to the skulls. Comments on a few of the long bones which exhibit gross abnormalities are given in Table VIII, and the remaining parts of the skeletons have not been examined. The Neolithic skeleton of a young man (Q 1) was extensively mutilated, and otherwise it is not remarkable. The skull-cap of the same period is also that of a young man. The Saxon skeleton (Q) is male but the age at death cannot be estimated as the skull is missing. The following remarks refer to the remaining specimens, which are all of Iron Age or Romano-British date.

Estimates of the ages at death of the adults were obtained by noting the conditions of the coronal, sagittal and lambdoid sutures, and the ectocranial aspects of these give the following frequencies:

		All sutures open	Sutures beginning to close or partly closed	All sutures closed	Totals
♂	Belgic War Cemetery series	5 (28%)	12 (67%)	1 (6%)	18
	Others	3 (21%)	9 (64%)	2 (14%)	14
	Total series	8 (25%)	21 (66%)	3 (9%)	32
♀	Belgic War Cemetery series	3 (43%)	4 (57%)	—	7
	Others	7 (37%)	10 (53%)	2 (11%)	19
	Total series	10 (38%)	14 (54%)	2 (8%)	26

It is probable that all the people buried in the Belgic War Cemetery were massacred. The age constitutions of the short adult series are very similar, however, to those of the series made up by other people interred at Maiden Castle, who probably died from natural causes. The massacre must have been carried out without regard to age, though there appears to have been some sex discrimination, since twice as many men as women were excavated from the cemetery (see Table I). The percentage frequencies for the total series show the usual sex differences, due to the fact that the sutures tend to close at an earlier age in males than in females. Comparison with values obtained in the same way for other series (see Risdon, 1939, p. 107) shows that the Iron Age men died at a rather younger age, on the average, than those interred in English cemeteries of a later date, but that the women are not distinguished in the same way. In considering the frequencies of anomalies, the total series from Maiden Castle is referred to below.

The sagittal suture was normally the first to close, followed by the coronal and then the lambdoid. Only one specimen (P 23, male) is definitely anomalous with regard to the condition of the principal calvarial sutures. Its sagittal suture is obliterated, the coronal is beginning to close and the lambdoid is open. This specimen shows no apparent sign of distortion and its cephalic index (77·4) is above the average, so it may be assumed that the sagittal suture was not obliterated before maturity was reached. Of thirty-three male frontal bones only one (P ii, an isolated bone for which no measurements are given) is metopic, and there is only one female metopic specimen out of a total of twenty-seven frontal bones. Among Western European crania metopic specimens are usually found with a frequency of about 10%, but the Iron Age series are so short that no significance can be attached to the fact that the frequencies for them are exceptionally low. There are no examples of interparietal bones, or of an *os*

épactal, among the thirty-two male and twenty-six female occipital bones. A much rarer anomaly is exhibited by a male cranium (P 7A). This is a trace of a suture (33 mm. long) extending across the posterior part of the right parietal bone: its position can be seen from Plate IIB. One male skull (P 27) shows the occipito-mastoid suture obliterated prematurely on both sides. There is one case of fronto-temporal articulation on both sides among the males, and one of the pterion in *K* on both sides among the females. One female specimen has both malar bones divided by horizontal sutures. A male skull (O 3, Plate IIIB) has the basi-occipital broken exposing the extension into it of the sphenoidal air sinuses, the condition exhibited by the Swanscombe skull. There is one large cavity and the extremity of a small one to the right of it.

Wounds that had healed completely during life were noted on the cranial vaults of three of the male and two of the female skulls. None of these was severe and a female specimen (T 22) has a depression on the right temple which was more serious than the other injuries. An adolescent female skull (P 20, Plate II A) has the right malar bone deformed, probably as the result of a wound. By far the most serious injuries of traumatic origin are shown by the long bones of one of the males (T 5). His right elbow was shattered, involving complete separation of the proximal extremity of the ulna, his left radius was fractured near the wrist, and his fractured left fibula became fused to the tibia. Healed fractures of the long bones were only found in the case of two other skeletons, viz. P 23 (*male, right ulna*) and P 36 (*female, left fibula*). *Pathological conditions* of bones not of traumatic or dental origin were only noted in the case of one male skull (T 9) which has a swelling on the right maxilla below the orbit which was probably caused by a tumour, of a skeleton of the same sex (P 25) which has osseous deposits on the muscular ridges of the left femur and fibula, and of a female (T 12) which has the left humerus markedly deformed, probably as the result of an inflammatory condition of the surrounding soft tissues. We are indebted to Dr A. M. El Batrawi for help in interpreting the conditions of these and other specimens in the series.

The teeth of the individuals interred at Maiden Castle are by no means well preserved. The following frequencies are found for the total "Iron Age" series:

	Upper jaw		Lower jaw	
	♂	♀	♂	♀
No teeth lost before death	12	8	22	13
One or more teeth lost before death	16	15	10	12

In the case of excavated series of skulls which are not of recent date, it is customary to find for either sex that more than half the specimens had lost no

teeth from the upper jaw before death. Abscess cavities were found in several of the jaws. A female specimen (T 21, Plate III A) has two symmetrically disposed cyst cavities in the buccal surfaces of the maxillae at the roots of the second premolars. There are examples of crowded front teeth. A male mandible (P 23) has a milk molar retained in the position of the left second premolar, which had not appeared, and a female bone (P 26) appears to have had only three incisors erupted. The upper jaw of a female skull (T 28, Plate III D) shows the canines missing and the first premolars rotated. A female mandible (P 14, Plate III E) has the left third molar impacted but almost fully erupted: there are also distinct swellings on the inner alveolar margin extending from the first premolar to the first molar on the left side, and from the second premolar to the first molar on the right. Several of the *Sinanthropus* mandibles exhibit this condition (*torus mandibularis*), and it occurs more frequently, and to a more marked extent, in some modern races of man (particularly in Eskimos), but is seldom found in European series. Dental anomalies appear to be exceptionally frequent in the short series of skulls from Maiden Castle.

In general appearance the specimens show considerable individual differences, but nearly all might be considered quite unexceptional if found in any British collection of later date than the Bronze Age. There are few examples, however, of the markedly retreating frontal bones which characterize the seventeenth-century London skulls, particularly those from Farringdon Street. The variation exhibited appears to be no greater than that expected for a community of intermarrying people, except for the fact that one female skull from the Belgic War Cemetery (P 36, Plate II C, D) stands apart from the others on account of its aberrant form. The facial skeleton has an unusual premaxillary height, though it is not prognathous; the nasal index is high but not extreme, and the nasal bridge is broad and depressed. The cephalic index of this specimen (87·0) is easily the highest for the series, and the individual was decidedly short (1451 mm. = 4 ft. 9¼ in.), though taller than two other women interred in the Belgic War Cemetery. It is possible that the skeleton P 36 is that of an alien in Western Europe, but it appears to be more probable that its peculiarities are of individual rather than racial significance. Its measurements were included in computing averages.

3. Metrical comparisons of the Iron Age and Romano-British crania

Excluding the two Neolithic specimens, there are 30 adult male and 26 adult female crania sufficiently complete to give measurements, though most of these are defective to some extent. Individual readings are in the appended Table VII. In spite of the small numbers, it appeared worth while computing separate means of the more important characters for (*a*) the Belgic War Cemetery series, and

(*b*) all the other specimens of Iron Age or Roman date. The latter is a miscellaneous collection made up by individuals belonging to the following groups:

	Male	Female		Male	Female
Iron Age *A*	1	—	Romano-British	3	3
„ *B*	1	7	Iron Age or Romano-British	2	1
„ *C*	6	8			

The pooling of this material, and treatment of it as if all the specimens represented a single racially homogeneous population, is obviously a *pis aller*. The means found for the composite group (Table III) are actually very close to those found for the Belgic War Cemetery series, which is almost certainly made up by individuals who belonged to a single community. On the supposition that the variabilities of the groups were of the usual order, all the differences between the two sets of means are quite insignificant. Both types have surprisingly large basio-bregmatic heights, and hence unusual indices ($100\,H'/L$ and $100\,B/H'$) involving this diameter. Otherwise, they appear to have no features which can be considered at all peculiar in English series of post-Bronze Age date.

This comparison provides some justification for combining the two subgroups, to give a single series which may be supposed to represent the population of Maiden Castle from Iron Age *A* to Roman times, though it is clear that the evidence available is quite inadequate to demonstrate that the racial constitution of the population remained stable throughout the period. The following distributions are found for the total sample.

Cephalic index (central values)	67	69	71	73	75	77	79	81	83	85	87	Totals
♂	1	1	2	5	6	7	—	1	—	—	—	23
♀	—	—	2	4	9	4	—	2	—	—	1	22

Height-length index (central values)	68·5	69·5	70·5	71·5	72·5	73·5	74·5	75·5	76·5	77·5	78·5	79·5	Totals
♂	—	3	1	4	3·5	3·5	2	2	1	—	—	1	21
♀	2	1	1	1	2	4	6	1	2·5	0·5	1	—	22

The highest cephalic index (87·0) is for a female specimen (P 36), in the Belgic War Cemetery series, which is also remarkable on account of the fact that its facial skeleton is of an unusual form. The highest height-length index is for a

TABLE III

Mean measurements of series of Iron Age and Romano-British crania from Maiden Castle

	Male			Female		
	Belgic War Cemetery	Others	Total	Belgic War Cemetery	Others	Total
L	187·7 (14)	189·9 (9)	188·6 (23)	179·0 (7)	180·5 (19)	180·1 (26)
B	141·4 (14)	140·0 (10)	140·8 (24)	135·7 (6)	135·8 (16)	135·8 (22)
H'	137·1 (14)	136·9 (8)	137·1 (22)	134·4 (7)	131·7 (15)	132·6 (22)
B'	97·2 (14)	96·9 (12)	97·1 (26)	92·5 (7)	94·1 (19)	93·7 (26)
S	384·5 (11)	381·9 (7)	383·5 (18)	371·3 (6)	365·6 (10)	367·7 (16)
U	528·6 (12)	530·9 (8)	529·5 (20)	504·0 (6)	508·1 (15)	507·0 (21)
fml	37·8 (13)	37·3 (8)	37·6 (21)	35·9 (5)	35·3 (11)	35·5 (16)
fmb	31·6 (12)	31·2 (8)	31·5 (20)	28·2 (5)	28·9 (10)	28·6 (15)
LB	102·5 (13)	102·2 (6)	102·4 (19)	96·7 (7)	99·9 (12)	98·7 (19)
$G'H$	73·3 (8)	71·3 (6)	72·4 (14)	68·1 (6)	67·2 (9)	67·6 (15)
NH, L	52·4 (9)	51·2 (7)	51·9 (16)	46·7 (6)	49·1 (12)	48·3 (18)
NB	25·0 (8)	25·1 (8)	25·0 (16)	22·4 (5)	24·3 (8)	23·5 (13)
$O_1 L$	43·6 (9)	41·2 (6)	42·6 (15)	41·3 (5)	41·3 (9)	41·3 (14)
$O_2 L$	33·2 (9)	32·3 (6)	32·9 (15)	32·5 (5)	31·9 (9)	32·1 (14)
$100\ B/L$	75·4 (14)	73·5 (9)	74·6 (23)	76·6 (6)	75·5 (16)	75·8 (22)
$100\ H'/L$	73·4 (13)	72·4 (8)	73·0 (21)	75·2 (7)	72·9 (15)	73·6 (22)
$100\ B/H'$	103·3 (13)	102·7 (8)	103·1 (21)	101·6 (6)	102·5 (13)	102·2 (19)
$100\ fmb/fml$	83·5 (12)	83·7 (8)	83·6 (20)	79·2 (5)	82·9 (9)	81·5 (14)
$100\ NB/NH, L$	48·6 (7)	50·2 (6)	49·4 (13)	48·1 (5)	50·0 (8)	49·3 (13)
$100\ O_2/O_1, L$	76·3 (9)	78·5 (6)	77·2 (15)	78·8 (5)	77·2 (9)	77·8 (14)
$P\angle$	86°·7 (8)	86°·0 (4)	86°·5 (12)	89°·0 (3)	83°·9 (9)	85°·2 (12)

Total series

	Male	Female
$\beta Q'$	316·7 (20)	305·3 (23)
S_1'	114·2 (24)	110·2 (22)
S_2'	116·5 (22)	112·2 (21)
S_3'	99·9 (22)	97·0 (17)
S_1	132·0 (22)	126·6 (22)
S_2	129·9 (21)	125·3 (21)
S_3	122·3 (22)	116·0 (17)
C*	1514·7 (19)	1380·7 (19)

	Male	Female
B''	122·4 (17)	114·7 (16)
Biast. B	114·3 (19)	107·2 (14)
GL	93·4 (14)	90·3 (11)
GB	95·2 (18)	89·4 (14)
J	133·6 (11)	123·2 (11)
G_1'	45·6 (19)	44·0 (12)
G_2	40·7 (16)	39·0 (11)
SS	4·7 (16)	4·3 (14)

	Male	Female
SC	10·1 (16)	9·3 (15)
Oc.I.	58·7 (22)	60·5 (17)
$100\ G'H/GB$	76·3 (13)	73·5 (11)
$100\ G_2/G_1'$	88·7 (12)	87·4 (7)
$100\ SS/SC$	47·6 (16)	47·6 (14)
$N\angle$	61°·0 (14)	63°·2 (11)
$A\angle$	76°·3 (14)	75°·2 (11)
$B\angle$	42°·6 (14)	41°·6 (11)

* Reconstructed from L, B and H'.

male cranium (P 7 A) in the same series. In spite of these outlying cases, the distributions for the two characters in question, and those for the other characters, provide no clear evidence of racial heterogeneity, though it cannot be inferred from them that the total population considered was racially homogeneous.

If the supposition that all the skulls of Iron Age and Romano-British date from Maiden Castle represent a single population be accepted, the series available is still not long enough to give reliable racial comparisons of a statistical kind. A short series is liable to indicate an absence of differentiation from other series when one of an adequate length from the same population will provide evidence of distinction, and it will also be liable to give a misleading estimate of divergence from other types when distinction is clearly indicated. Large enough samples must be demanded, in particular, when making comparisons between closely related populations, such as the group made up by all the prevailing populations of England from the end of the Bronze Age to modern times. In spite of its restricted size, coefficients of racial likeness were computed for male means between the total "Iron Age" series from Maiden Castle and four others, viz.:

(*a*) The Anglo-Saxon made up principally by skulls preserved in London Museums (Morant, 1926).

(*b*) The so-called British Iron Age (Morant, 1926) made up by skulls from the south of Scotland and various parts of England. Several of the specimens are known to be of Romano-Britons, and some of the others are not dated satisfactorily, so the series is of little value.

(*c*) The seventeenth-century series from a plague in Whitechapel (Macdonell, 1904), the revised means given by Hooke (1926) being used.

(*d*) The seventeenth-century series from the cemetery in Farringdon Street (Hooke, 1926).

The following coefficients of racial likeness are found for the total "Iron Age" series from Maiden Castle and these four, the numbers in brackets following the crude values being the numbers of characters on which they are based, and the numbers in brackets following the names of the series being the average numbers of skulls on which the means used are based ($\bar{n}$'s)*:

	Crude C.R.L.	Reduced C.R.L.
Maiden Castle (17·7) and Anglo-Saxon (34·1)	$-0·28 \pm 0·18$ (29)	—
„ (18·5) and British Iron Age (55·4)	$2·98 \pm 0·21$ (20)	$10·73 \pm 0·77$
„ (17·7) and Whitechapel (90·7)	$2·79 \pm 0·17$ (30)	$9·43 \pm 0·59$
„ (17·8) and Farringdon Street (96·7)	$4·29 \pm 0·17$ (30)	$14·27 \pm 0·58$

As far as can be told from the scanty evidence, the Maiden Castle and Anglo-Saxon samples might represent different sections of the same population, while the former series is clearly differentiated from the other three. Previous comparisons have shown that the British Iron Age and the two seventeenth-century London types are very similar, while the Anglo-Saxon stands apart from that cluster. These relationships suggested that the Londoners were descended primarily from the pre-Saxon rather than from the Anglo-Saxon population, but the new evidence does not support this view.

* The standard deviations of the Farringdon Street series were used in computing the coefficients.

For characters considered singly, there are very few significant differences between the means for the short Maiden Castle series and those for the three later series. All the differences from the Anglo-Saxon values are quite insignificant, the highest of the 29 α's being 3·8. The only α's greater than 10 are for the nasal angle ($N\angle$, 15·0) and basio-bregmatic height (H', 11·8) in comparisons with the British Iron Age means, for H' (39·7), 100 H'/L (32·7) and 100 B/H' (28·7) in comparisons with the Farringdon Street means, and for H' (18·9), 100 H'/L (15·3), $N\angle$ (15·2) and 100 B/H' (10·5) in comparisons with the Whitechapel means. The Maiden Castle skulls are markedly orthognathous judging from the nasal angle, but this measurement is only available for 14 male skulls and little stress can be laid on the peculiarity. Otherwise the type is only distinguished by a calvarial height which is large both absolutely and also relative to the length and breadth of the brain-box. It is known that the Anglo-Saxon is distinguished in precisely the same way from the other types considered.

Table IV gives male means of the three principal diameters and the three indices derived from them for the series referred to above and the following:

(*e*) Three series of Romano-British skulls described by Buxton (1935), the six measurements in question being the only ones available for these.

(*f*) An Anglo-Saxon series from Burwell, Cambridgeshire (Brash *et al.* 1935).

(*g*) A third seventeenth-century London series from a burial-pit at Moorfields (Macdonell, 1906).

(*h*) A modern series of Lowland Scottish skulls measured by Turner (1903) and compiled in the way described by Hooke (1926, pp. 22 and 38).

(*i*) A modern series from Glasgow (Young, 1931).

All the series included in the table are believed to represent populations which prevailed in England and the south of Scotland at certain periods from the beginning of the Iron Age to modern times, and the minority populations for which there is craniological evidence are omitted. The length, breadth and cephalic index are seen to be remarkably constant throughout, while the basio-bregmatic height and the two indices involving this diameter show larger differences. The greatest heights, highest height-length and lowest breadth-height indices are for the two series from Maiden Castle, and they thus show a remarkable concordance in spite of their very restricted sizes. In these respects one of the Romano-British series (the Brigantes) and the two Anglo-Saxon come next, and the others follow, with one other Romano-British (the Dobuni) and two of the seventeenth-century London series at the lower end of the scale. Little significance can be attached to this order, however, as some of the series are short. There are no statistically significant differences, for example, between the means for the total series from Maiden Castle and those for the Romano-British Belgae. The populations represented in Table IV are all remarkably similar in type, and they must have been closely inter-related. Larger series

TABLE IV

Mean measurements of British series of male skulls

Iron Age and Romano-British series	Maiden Castle			"British Iron Age"	Romano-British		
	Belgic War Cemetery	Others	Total		Belgae	Dobuni	Brigantes
L	187·7 (14)	189·9 (9)	188·6 (23)	187·4 (61)	189·6 (40)	190·8 (85)	189·9 (57)
B	141·4 (14)	140·0 (10)	140·8 (24)	141·4 (102)	141·0 (41)	144·2 (77)	141·7 (57)
H'	137·1 (14)	136·9 (8)	137·1 (22)	132·9 (77)	134·2 (33)	132·5 (45)	136·8 (38)
$100\ B/L$	75·4 (14)	73·5 (9)	74·6 (23)	75·4 (61)	74·4 (40)	75·6 (71)	75·7 (51)
$100\ H'/L$	73·4 (13)	72·4 (8)	73·0 (21)	70·9 (61)	71·0 (33)	69·4 (43)	71·4 (34)
$100\ B/H'$	103·3 (13)	102·7 (8)	103·1 (21)	106·3 (77)	105·5 (33)	109·9 (41)	104·4 (33)

Other series	Anglo-Saxon		Seventeenth-century London			Modern Scottish	
	London museums	Burwell	Farringdon Street	White-chapel	Moorfields	Lowland	Glasgow
L	190·6 (58)	189·6 (45)	188·8 (139)	189·1 (137)	189·2 (44)	188·8 (54)	188·2 (524)
B	141·7 (103)	141·7 (45)	142·4 (141)	140·7 (135)	143·0 (46)	142·1 (54)	139·1 (524)
H'	136·0 (31)	136·3 (40)	129·7 (118)	132·0 (122)	129·8 (34)	133·6 (52)	132·9 (521)
$100\ B/L$	74·7 (52)	74·8 (45)	75·4 (132)	74·3 (131)	75·5 (42)	75·3 (54)	75·0 (524)
$100\ H'/L$	71·2 (25)	71·9 (40)	68·6 (115)	70·0 (120)	68·4 (31)	70·9 (52)	70·7 (521)
$100\ B/H'$	104·9 (61)	104·3 (40)	109·8 (117)	106·6 (122)	110·2 (34)	106·4 (52)	104·9 (521)

representing the Iron Age population are obviously required, but the data available clearly focus attention on mean differences between the absolute and relative magnitudes of the calvarial height. It can be seen from the distribution on p. 301 above that all the male skulls from Maiden Castle have height-length indices greater than 69·0. In Table IV there are two seventeenth-century London series with means for the index less than this value, and one of the Romano-British series has a mean of 69·4. These conditions indicate unusual separation of the distributions for different series, and there is no reason to suspect that any cranial characters other than the height and the height indices would distinguish the populations represented in Table IV as effectively.

4. MEASUREMENTS OF THE MANDIBLES

Measurements of the mandibles were taken in accordance with the revised technique given in *Biometrika* (Morant *et al.* 1936, Appendix), and readings for individual bones are not provided in the present paper. It has been found (Cleaver, 1937) that larger numbers of lower jaws than of crania are needed to

give any decisive racial comparisons, and hence the means for all the adult specimens of Iron Age and Romano-British date taken together are the only ones worth considering. They are given in Table V for totals of thirty-four male and twenty-five female bones, and values for Anglo-Saxon (Morant, 1926) and a seventeenth-century London (Cleaver, 1937) series are included for comparative purposes.

TABLE V

Mean measurements of English series of mandibles

	Male			Female		
	Anglo-Saxon	Maiden Castle: Iron Age	Farringdon Street, London: seventeenth century	Anglo-Saxon	Maiden Castle: Iron Age	Farringdon Street, London: seventeenth century
w_1	123·7 (25)	121·5 (19)	117·7 ± 0·53 (23)	116·6 (22)	117·5 (12)	110·8 ± 0·64 (30)
g_0g_0	100·4 (33)	99·5 (25)	97·7 ± 0·70 (40)	92·9 (35)	90·8 (18)	85·7 ± 0·55 (50)
c_rc_r	100·3 (27)	99·3 (20)	95·9 ± 0·68 (29)	93·2 (28)	95·9 (18)	91·7 ± 0·54 (35)
zz	45·3 (57)	44·8 (33)	43·9 ± 0·24 (40)	44·1 (50)	42·5 (24)	42·9 ± 0·25 (49)
c_yl	21·7 (38)	21·0 (27)	19·8 ± 0·20 (34)	19·1 (35)	20·0 (14)	18·0 ± 0·16 (45)
ml	107·2 (31)	105·3 (24)	104·1 ± 0·64 (34)	104·2 (45)	98·9 (16)	99·4 ± 0·61 (43)
c_pl	77·2 (42)	79·0 (27)	74·9 ± 0·40 (40)	74·6 (49)	70·5 (19)	69·7 ± 0·37 (50)
rb'	33·2 (61)	33·1 (33)	30·9 ± 0·28 (40)	31·0 (56)	30·4 (24)	28·3 ± 0·25 (50)
m_2p_1	28·1 (59)	28·4 (23)	28·2 ± 0·18 (22)	27·6 (57)	27·2 (19)	27·7 ± 0·22 (19)
h_1	33·1 (40)	34·7 (22)	30·9 ± 0·46 (12)	30·5 (31)	30·5 (17)	29·7 ± 0·36 (26)
m_2h	27·2 (51)	28·2 (20)	24·9 ± 0·45 (19)	24·4 (52)	26·6 (15)	23·6 ± 0·51 (12)
c_rh	65·7 (48)	70·6 (28)	64·8 ± 0·47 (40)	59·2 (47)	60·2 (25)	56·5 ± 0·47 (50)
rl	64·0 (45)	66·2 (26)	62·2 ± 0·40 (36)	59·1 (45)	57·3 (20)	53·5 ± 0·44 (43)
$M\angle$	120°·3 (47)	116°·5 (27)	121°·7 ± 0·60 (40)	122°·5 (49)	123°·1 (19)	127°·8 ± 0·59 (50)
$R\angle$	72°·0 (36)	77°·6 (24)	72°·0 ± 0·94 (37)	68°·2 (36)	71°·6 (18)	71°·2 ± 0·87 (48)
$C'\angle$	68°·2 (32)	69°·8 (18)	61°·8 ± 1·35 (16)	70°·0 (25)	70°·3 (14)	67°·3 ± 0·80 (27)
100 c_rh/ml	60·9 (27)	67·9 (22)	62·4 ± 0·66 (34)	58·3 (38)	60·5 (16)	56·8 ± 0·49 (43)
100 c_rc_r/ml	94·4 (15)	93·9 (17)	92·3 ± 1·03 (25)	91·7 (26)	97·1 (12)	92·4 ± 0·74 (30)
100 g_0g_0/c_pl	129·0 (32)	127·4 (25)	130·9 ± 1·23 (40)	126·2 (35)	129·2 (18)	123·3 ± 1·02 (50)
100 rb'/rl	51·5 (45)	50·8 (26)	50·0 ± 0·54 (36)	53·0 (43)	53·5 (19)	53·2 ± 0·66 (43)
100 g_0g_0/c_rc_r	99·3 (19)	102·1 (18)	102·9 ± 1·07 (29)	99·3 (23)	96·2 (14)	94·0 ± 0·86 (35)

None of these is long enough to give an adequate representation of the type for the population it represents, and probable errors are only available for the last. A few general comparisons are sufficiently suggestive, however, to be of interest. For both sexes the two earlier types are very similar in size, and for several characters both appear to be significantly larger than that of the Farringdon Street series. There is only one measurement of size which provides a clear exception to this rule—viz. the length of the dental arcade from second molar to first premolar (m_2p_1)—and it is the only measurement taken relating to the size of the teeth. Both Anglo-Saxon and Iron Age types appear to be distinguished from that of seventeenth-century Londoners on account of less protruding chins ($C'\angle$ greater), and the Iron Age appears to be distinguished

from the other two on account of a more outstanding coronoid process (judging from $R\angle$ and 100 $c_r h/ml$). Otherwise, there is no clear evidence of distinctions between the types.

Judging from all the characters, the Anglo-Saxon and Maiden Castle populations were just distinguished by features of their lower jaws, but the resemblance between them was closer than that between either and the later Londoners. The relations found favour the hypothesis that the mandibles of Englishmen, but not their teeth, became slightly smaller in historical times, but more data will be needed to substantiate it. If the hypothesis be accepted, then mandible measurements should not be used to estimate racial relationships unless allowance is made for secular changes in them within the same population.

5. MEASUREMENTS OF THE LENGTHS OF THE LONG BONES

Individual readings are given in Table VIII, and means of the adult series for the characters of greatest interest are in Table VI. The lengths were taken in the ways specified by Münter (1936), whose means for Anglo-Saxon skeletons are quoted. In spite of the small numbers, it was thought worth while computing separate means for (*a*) the Belgic War Cemetery series, and (*b*) for all the other specimens of Iron Age or Roman date from Maiden Castle.

For both sexes all the absolute measurements for these two series in Table VI are less than the corresponding values for Anglo-Saxons. A rough appreciation of the significance of the differences between the means can be obtained by supposing that the two Maiden Castle populations considered, and also that made up by combining them, exhibited the same variabilities as the total Anglo-Saxon population. This may appear to be a very arbitrary assumption, but in fact it is not unreasonable, since a close approach to equality in variation is usually found on comparing different subgroups with a total population, and also on comparing distinct populations.

By applying the Anglo-Saxon standard deviations, we reach the conclusion that there are no significant differences between the means of the absolute and indicial measurements in the case of the Belgic War Cemetery and the other series from the same site, and this is true for both sexes. On the same assumption with regard to variation, it is found that the pooled Maiden Castle means only show markedly significant differences from the Anglo-Saxon in the case of the length of the femur for males, the length of the humerus for males and females, and the radius-humerus index for males. The differences between the statures cannot be supposed clearly significant in the case of either sex, but it is safe to conclude from the evidence of both that the Iron Age inhabitants of Maiden Castle were shorter than Anglo-Saxons. The sex ratios for stature are almost identical, being 1·076 for the former and 1·073 for the latter series, and of the order usually found. In inches the estimates of height obtained are

TABLE VI

Means for lengths of the right femur, tibia, humerus and radius, and indices derived from these lengths, for Maiden Castle (Iron Age and Romano-British) and Anglo-Saxon series of adult skeletons

	Male			
	Other than Belgic War Cemetery	Belgic War Cemetery	All Maiden Castle	Anglo-Saxon
F. max.	437·6 (11)	443·2 (15)	440·8 (26)	463·3 ± 1·22 (153)
T. max.*	362·8 (11)	371·2 (18)	368·0 (29)	378·9 ± 1·46 (103)
H. max.	323·0 (12)	326·8 (17)	325·2 (29)	337·1 ± 1·15 (121)
R. max.	240·8 (10)	250·9 (15)	246·9 (25)	251·6 ± 1·12 (79)
100 T. obl./F. obl.	81·2 (10)	82·5 (14)	82·0 (24)	81·1 ± 0·17 (92)
100 H. max./F. obl.	73·9 (11)	74·3 (13)	74·1 (24)	73·5 ± 0·14 (100)
100 R. max./H. max.	75·4 (10)	77·2 (15)	76·4 (25)	74·6 ± 0·19 (62)
100 (H. max. + R. max.)/(F. obl. + T. max.)†	71·2 (9)	71·6 (12)	71·5 (21)	70·7 ± 0·16 (41)
Reconstructed stature	1640 (13)	1654 (19)	1649 (32)	1683 (161)‡
	Female			
F. max.	411·9 (13)	411·0 (8)	411·5 (21)	426·1 ± 2·05 (56)
T. max.*	336·4 (14)	342·7 (7)	338·5 (21)	350·4 ± 1·99 (44)
H. max.	300·8 (11)	296·4 (8)	298·9 (19)	312·5 ± 1·54 (47)
R. max.	223·0 (13)	215·1 (7)	220·3 (20)	227·6 ± 1·41 (34)
100 T. obl./F. obl.	81·5 (13)	81·2 (6)	81·4 (19)	80·8 ± 0·20 (38)
100 H. max./F. obl.	74·5 (8)	72·3 (7)	73·5 (15)	73·6 ± 0·25 (36)
100 R. max./H. max.	74·2 (9)	73·4 (7)	73·8 (16)	73·5 ± 0·22 (31)
100 (H. max. + R. max.)/(F. obl. + T. max.)†	70·8 (7)	69·7 (5)	70·3 (12)	70·6 ± 0·21 (21)
Reconstructed stature	1535 (17)	1527 (9)	1532 (26)	1568 (59)‡

* Maximum length of the tibia including the spine. † Maximum length of the tibia excluding the spine.
‡ The reconstructed statures for Anglo-Saxons were found from the mean lengths for different long bones, and the numbers of individuals given are the numbers of femora involved. The statures actually relate to rather larger numbers of skeletons.

5 ft. 6¼ in. for Anglo-Saxon and 5 ft. 5 in. for Iron Age men, and 5 ft. 1¾ in. for Anglo-Saxon and 5 ft. 0¼ in. for Iron Age women. The average stature of the general male adult population of England to-day is about 5 ft. 7½ in., and for different social classes means between about 5 ft. 5¾ in. and 5 ft. 9¼ in. are found. The Iron Age men were thus decidedly short compared with modern Englishmen, and the estimated stature for them is very close to the average found for all European populations to-day.

For the measurements considered, the only distinction between the two ancient populations other than that in size is found for the index expressing the length of the radius as a percentage of the length of the humerus in the case of

the male, but not in that of the female, sample. It might be suggested that the proportions of the upper and fore-arms of the men who lived at Maiden Castle were influenced on the right, or on both, sides by continual practice in the use of the sling, which is known to have been one of their principal weapons from the evidence of caches of stones. The following means, which relate only to paired indices, are of interest in this connexion:

	100 × Radius max./Humerus max.				100 × (H. max. + R. max.)/ (F. obl. + T. max. ex spine)			
	Maiden Castle		Anglo-Saxon		Maiden Castle		Anglo-Saxon	
	Male	Female	Male	Female	Male	Female	Male	Female
R.	76·3 (18)	73·6 (14)	74·6 (25)	74·0 (14)	71·5 (11)	70·3 (12)	70·9 (21)	70·4 (9)
L.	78·4 (18)	75·0 (14)	75·7 (25)	74·6 (14)	70·8 (11)	68·9 (12)	70·0 (21)	69·6 (9)
R.-L.	−2·1	−1·4	−1·1	−0·6	+0·7	+1·4	+0·9	+0·8

All the series are very restricted in size, but there is complete agreement in the signs of the side and sex differences found. The most significant difference between the two populations appears to be that for the male radius-humerus index on the left side. This might suggest that the Maiden Castle men were left-handed slingers, but their intermembral indices are so close to the Anglo-Saxon values that it seems unsafe to accept the hypothesis that the lengths of their arms were affected by use. More abundant material might tell definitely for or against such a view, and a detailed anatomical examination of the arm bones of the Maiden Castle and modern series would also be relevant to the question.

6. SUMMARY AND CONCLUSIONS

The material studied consists of the imperfect skeletal remains of eighty-three individuals, who are distributed in Table I according to periods, sexes, and ages at death. This paper is concerned chiefly with the customary measurements of the crania and mandibles, and with the lengths of the long bones. Comparisons of various kinds are made between:

A, the Belgic War Cemetery series, representing the defenders of the Castle who were massacred by Roman invaders in A.D. 43, and

B, a composite series made up by all the other specimens of Iron Age and Roman date.

There is no clear distinction between the distributions of the ages at deaths of the adults forming the two short series, but for both the men died at a rather younger age, on the average, than those interred in English cemeteries of later

date, while the females are not distinguished in the same way. Remarks on cranial anomalies are provided. The teeth of the inhabitants of Maiden Castle were not well preserved.

Judging from their appearance, nearly all the crania in series *A* and *B* would be considered quite unexceptional if found in any British collection of later date than the Bronze Age, but there are few examples of the markedly retreating frontal bone which characterizes seventeenth-century Londoners. There is agreement between the male and female samples and this is confirmed by measurements of the skulls and long bones.

There are no statistically significant differences between the mean cranial (Table III) and long bone (Table VI) measurements of series *A* and *B*. Both are characterized by a large calvarial height, and the estimates of stature they give are about 1 in. less than the Anglo-Saxon values. The type of the total series $(A+B)$ of crania is found to be indistinguishable from the Anglo-Saxon, while both are differentiated from the types of seventeenth-century London series. This conflicts with the conclusion reached previously, from very inadequate data, for Iron Age and Romano-British crania, to the effect that the Iron Age and recent types are very similar while the Anglo-Saxon stands apart from both. More abundant material representing the Iron Age and Romano-British populations will be required to disclose the relationships of these closely allied groups. The cephalic index is practically constant for them, and the cranial types are distinguished most clearly by differences in the absolute and relative magnitude of the calvarial height (Table IV).

Measurements of the mandibles make a slight distinction between the Maiden Castle and Anglo-Saxon series, while both types are larger than that of seventeenth-century Londoners. The Iron Age men, but not the women, are distinguished from Anglo-Saxons by having a larger radius-humerus index. It is not clear that this difference is due to the fact that the men at Maiden Castle were slingers.

APPENDIX

Tables (VII and VIII) of individual measurements and remarks

The letters denoting periods (or groups) given in the third columns of the tables are: N. = Neolithic, I.A. = Iron Age (*A*, *B* or *C*), S. = Saxon, R.-B. = Romano-British, B. = Belgic, B.W.C. = Belgic War Cemetery. The letters denoting cranial measurements are those used in all craniometric papers in *Biometrika*. A list of them is given in the present volume, p. 162, but this does not include B'' = maximum frontal breadth (Martin, No. 10), and *Biast. B.* = biasterionic breadth (M. 12). Owing to their fragility, the capacities of the crania could not be determined by any direct method. The estimates given in Table VII were obtained by applying the reconstruction formulae involving L, B and H' given by Hooke (1926, p. 33). The reconstructed statures in Table VIII were obtained by applying the formulae given by Pearson (1898), using as many as possible of the bones involved in each case. The lengths of the long bones were determined in the ways adopted by Münter (1936). All the readings given in the two tables, whether queried or not, can be considered close approximations to the true values, with the possible exception of the few measurements of

the Neolithic skeletons enclosed in square brackets. These are more uncertain than the others, but they are given because the specimens are of particular interest.

Unless otherwise stated in the remarks given in Table VII, the cranium and mandible are complete, or almost complete, with both dental arcades intact and no teeth lost from either jaw before death, and the coronal, sagittal and lambdoid sutures are open. Absence of a third molar denotes that the tooth had never erupted, as far as can be told. Two male skulls are not in the tables, since no measurements can be given for them, but they were included in determining the frequencies of qualitative features. These are:

P 24 (B.W.C.). Sagittal suture closed, coronal and lambdoid beginning to close. Upper jaw incomplete and some teeth lost before death, apparently only 1 incisor R.; 2 teeth lost from mandible and M 3's absent, abscess cavities at roots of M 1's.

R 2 (? I.A.). Sagittal suture closing. Upper jaw missing. Superficial wound on frontal bone.

LIST OF PLATES

Plate I. A typical male skull (P 30) from the Belgic War Cemetery. This specimen has a cephalic index which is very close to the mean for the series, but its height-length index (70·9) is below the average (73·4) and it is peculiarly orthognathous ($P\angle = 90°$). The lower incisors are crowded.

Plate II. Exceptional skulls from the Belgic War Cemetery.
- A. A female adolescent cranium (P 20) with injury to the right malar bone.
- B. A male cranium (P 7 A) with a trace of a suture on the right parietal bone.
- C and D. A female skull (P 36) with an exceptional type of facial skeleton and high cephalic index (87·0).

Plate III. Skulls from Maiden Castle with dental and other anomalies.
- A. A female cranium (T 21) with cyst cavities at the roots of the second premolars.
- B. The broken basi-occipital part of a male cranium (O 3) showing the extension into it of sphenoidal air sinuses.
- C. The palate of a male cranium (P 7) with a large anterior palatine foramen.
- D. The palate of a female cranium (T 28) showing absence of the canines and rotation of the first premolars.
- E. A female mandible (P 14) with the left third molar impacted and mandibular torus.

REFERENCES

BRASH, J. C., LAYARD, DORIS & YOUNG, M. (1935). "The Anglo-Saxon skulls from Bidford-on-Avon, Warwickshire, and Burwell, Cambridgeshire." *Biometrika*, **27**, 373–407.

BUXTON, L. H. D. (1935). "The racial affinities of the Romano-Britons." *J. Rom. Stud.* **25**, 35–50.

CLEAVER, F. H. (1937). "A contribution to the biometric study of the human mandible." *Biometrika*, **29**, 80–112.

GOODMAN, C. N. & MORANT, G. M. (MS.). "The human remains from Maiden Castle."

HOOKE, BEATRIX G. E. (1926). "A third study of the English skull with special reference to the Farringdon Street crania." *Biometrika*, **18**, 1–55.

MACDONELL, W. R. (1904). "A study of the variation and correlation of the human skull, with special reference to English crania." *Biometrika*, **3**, 191–244.

—— (1906). "A second study of the English skull, with special reference to Moorfields crania." *Biometrika*, **5**, 86–104.

MORANT, G. M. (1926). "A first study of the craniology of England and Scotland from Neolithic to early historic times, with special reference to the Anglo-Saxon skulls in London museums." *Biometrika*, **18**, 56–98.

MORANT, G. M., COLLETT, MARGOT & ADYANTHĀYA, N. K. (1936). "A biometric study of the human mandible." *Biometrika*, **28**, 84–122.

Münter, H. (1936). "A study of the lengths of the long bones of the arms and legs in man, with special reference to Anglo-Saxon skeletons." *Biometrika*, **28**, 258–94.

Pearson, K. (1898). "On the reconstruction of the stature of prehistoric races." *Philos. Trans.* A, **192**, 169–244.

Risdon, D. L. (1939). "A study of the cranial and other human remains from Palestine excavated at Tell Duweir (Lachish) by the Wellcome-Marston Archaeological Research Expedition." *Biometrika*, **31**, 99–166.

Turner, Sir William (1903). "A contribution to the craniology of the people of Scotland." *Trans. Roy. Soc. Edinb.* **40**, 547–613.

Wheeler, R. E. M. [Unpublished report on excavations at Maiden Castle.]

Young, M. (1931). "The West Scottish skull and its affinities." *Biometrika*, **23**, 10–22.

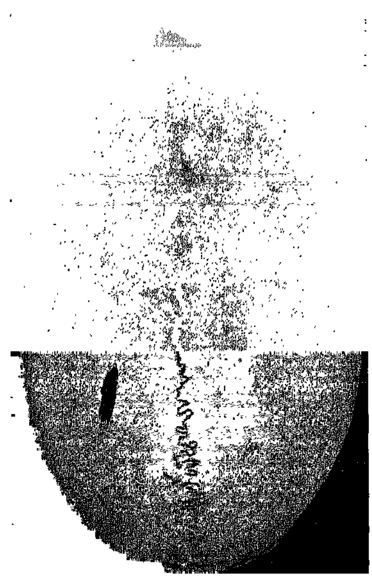
TABLE VII

M*ents of crania from Maiden Castle*

	G_1'	G_1	S	C	$100\frac{B}{L}$	$100\frac{H'}{L}$	$100\frac{B}{H'}$	Oc.I.	$100\frac{fmb}{fml}$	$100\frac{G'H}{GB}$	$100\frac{NB}{NH}$
	—	—	-	—	[71·8]	—	—	—	—	—	—
	—	—	-	—	[71·3]	—	—	58·7?	—	—	-
	—	—	-	—	73·0	74·3?	98·2?	—	—	—	-
	—	41?	—	—	—	—	—	—	82·3	—	
	43·9	—	-	—	—	—	—	56·2?	80·2	—	
	43·1	38	-	—	67·0?	—	—	—	—	—	
	42·0	—	5	1564	75·4	73·0	103·3	64·6	90·7	74·1	
9	45·5	40	9	1574	72·0	72·5	99·3	60·9	87·3	—	
3	41·1	—	3	1317	73·9	75·9	97·4	59·2	76·6	76·1	
	—	—	-	1514?	75·8?	73·7?	102·9	58·8?	—	—	
	—	40	-	—	—	—	—	61·0?	72·6?	—	
5	49·2	—	5	1551?	69·6?	72·7?	95·7?	58·2	95·1	72·7	
	—	43	-	1500	73·4?	72·9?	100·7?	56·8?	93·0?	—	
5	49·3	—	4	1688	76·2	79·1	96·3	61·1?	87·7	75·8	
7	—	—	2	1473	77·7	69·6?	111·6?	56·4	77·2	73·3	
	49·3	40	-	1563?	77·8?	73·3?	106·2?	60·9?	—	—	
	—	—	-	—	—	—	—	58·2?	—	—	
7	45·5	40	5	1496	77·4	76·6	101·1	58·6	82·4	80·	
	43·5?	39	-	—	—	—	—	—	—	-	
	—	—	-	—	71·6?	—	—	—	—		
1	48·0	38	5	1504	77·4	73·9	104·8	58·4	79·8		
	—	—	-	—	74·3?	74·3?	100·0?	—	—		
6	46·7	—	2	1651	76·5	71·6	106·9	56·2	87·7		
	—	39	-	1391	80·3	71·9	111·7	58·6	77·7		
1	—	—	3	1386	76·4	75·3	101·5	57·5	83·5		
3	43·5	40	8	1545	75·1	70·9	105·9	56·0	86·2		
4	45·2	41	1	1554?	72·0?	71·5	100·7?	59·0	79·6		
4	46·2	40	6	1453	75·7	69·3	109·3	57·3	84·8		
6	47·2	43	1	1522	75·3	71·2	105·7	62·4	88·6		
0	46·3	45	1	1534	73·1	69·4	105·2	55·9?	79·4		
0	47·1	39	5	—	—	—	—	—	—		
-	44·0?	—	-	—	—	—	—	—	—		
4	40·6?	—	1	1347	75·0	76·7	97·8	63·0	75·2		
9	—	—	6	1475	77·3	74·2	104·1	60·5	—		
8	47·2	36	7	1428?	73·3	74·1?	98·9?	—	—		
-	—	—	2	1463	74·5	76·9	96·8	61·9	88		
5	45·0	38	1	1173	75·1	73·4	102·4	63·4	8		
3	47·6	41	8	1391	75·0	74·4	100·7	59·4	8		
-	45·2	—	—	—	—	75·6	—	61·6			
-	—	—	—	1419	71·6	73·5	97·4	60·0			
8	—	—	3	1486	77·9	73·0	106·7	58·8?			
-	39·5?	35	—	1234	74·7	71·8	104·0	59·6			
-	—	—	—	1378?	77·2	69·2?	111·5?	—			
·7	—	37	3	—	—	68·9	—	—			
·8?	46·4	44	—	1479?	71·9?	68·8?	104·5?	59·2?			
-	—	40	3	—	74·2	—	—	—			
-	—	—	—	—	80·7	—	—	—			
-	41·0	—	—	1360	73·1	74·4	98·1	61·c			
·2?	41·6	33	7	—	—	74·2	—	—			
·7	—	—	4	1451?	74·1?	73·0	101·5?	58			
·7	—	40	—	1500?	72·3?	73·9	97·8?	5			
-	—	—	—	1197	75·4	74·9	100·8	5			
·8	48·7		—	1408	87·0	78·7	110·5				
·0	44·5		·0	1351?	77·6?	77·0?	100·7?				
-	—		—	—	—	—	—				
-	—	4	—	1397?	73·7?	70·7?	104·2?				
-	—		2·8	—	80·8	—	—				
·0	40·8		3·6	1297	74·4	72·1	103·				
-	—		—	1458?	83·7?	73·5?	113·				
-	41·9?		—	1477	78·0	77·4	100				

A

B



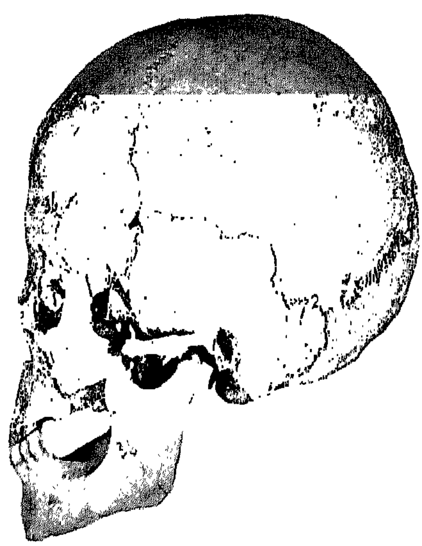
C

A typical male skull (P 30) from the Belgic War Cemetery.

A. P 20, female. Injury to right malar bone.

B. P 7 A, male. Trace of suture on right parietal bone.

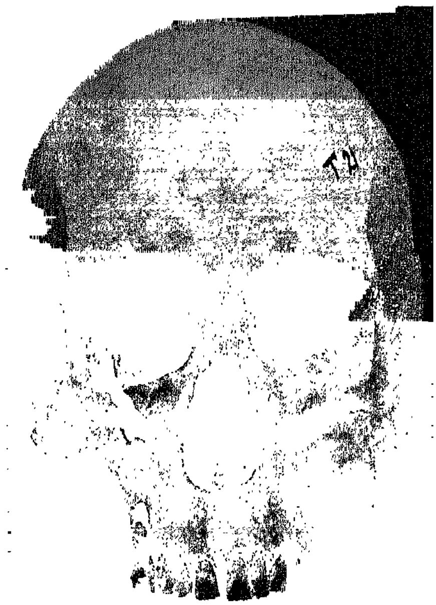
C. P 36, female. Exceptional type of facial skeleton and high cephalic index (87·0).



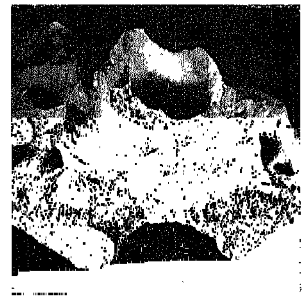
D. P 36, female. Profile view of the same skull (C).

Exceptional skulls from the Belgic War Cemetery.

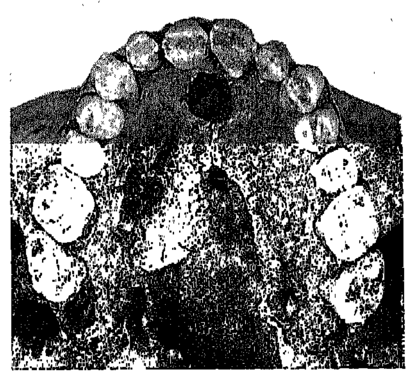
Goodman and Morant: *Human Remains from Maiden Castle*



A. T 21, female. Cyst cavities at roots of second premolars.

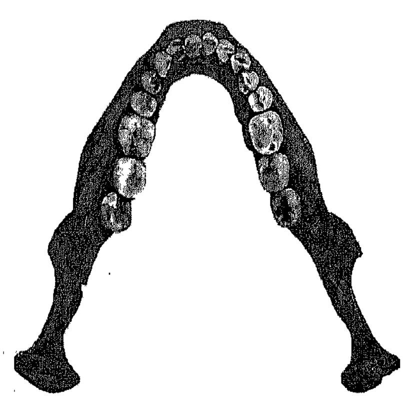


B. O 3, male. Broken basi-occipital showing the extension into it of sphenoidal air sinuses.



C. P 7, male. Large anterior palatine foramen.

D. T 28, female. Absence of canines and first premolars rotated.



E. P 14, female. Left third molar impacted and mandibular torus.

Skulls from Maiden Castle with dental and other anomalies.

HOMOGENEITY OF RESULTS IN TESTING SAMPLES FROM POISSON SERIES

WITH AN APPLICATION TO TESTING CLOVER SEED FOR DODDER

By J. PRZYBOROWSKI and H. WILENSKI
Plant Breeding and Agricultural Experimentation Department,
University of Krakow, Poland

1. Introductory

In our previous paper, "Statistical principles of routine work in testing clover seed for dodder" (Przyborowski & Wilenski, 1935), we justified the assumption that the number of dodder seeds in samples of clover does follow the Poisson Law:

$$p(x) = e^{-m}\frac{m^x}{x!} \quad (x = 0, 1, 2, \ldots). \tag{1}$$

Thus we constructed rules which should be followed in sampling problems where that Law holds good. We analysed particularly the application of the rules to the routine work in testing clover seed for dodder.

The purpose of this paper is to present some theoretical considerations relating to the problem of homogeneity of results in testing samples drawn from Poisson series which we have found arising in the course of our work on testing clover seed.

Let m_1 and m_2 denote the parameters of the Poisson distributions of the variables x_1 and x_2; the problem to be considered consists in testing the hypothesis $H_0(m_1 = m_2 = m)$, that the values x_1 and x_2 were obtained in sampling from Poisson series, the parameters of which have a common but unspecified value, say m.

The method of testing statistical hypotheses as developed by Neyman & Pearson (1933) consists in selecting a rule of rejecting the hypothesis in question, whenever the sample point E (the co-ordinates of which in the n-dimensional sample space W are the data of observation $x_1, x_2, \ldots, x_n$) lies within the so-called critical region, say w, of the sample space W. The probability, $P\{E\epsilon w \mid H_0\} = \alpha$, of rejecting the hypothesis tested, when it is true, is called the size of the corresponding critical region w on which the test is based. The errors thus committed when rejecting the true hypothesis H_0, are called the errors of the first kind. The probability $P\{E\epsilon w \mid H'\}$ of rejecting the hypothesis H_0 when an alternative H' is true has been termed the power of the test with regard to H'. $P\{E\epsilon w \mid H'\}$ considered as a function of H', where H' is any hypothesis belonging to the class Ω of alternatives to H_0, is called the power function of the test. The error committed when we fail to reject H_0 although

The hypothesis to be tested, H_0, is that $m_1 = m_2$ or that $\rho = \frac{1}{2}$; we shall suppose that the admissible alternatives to H_0 are both that $m_1 > m_2$ and $m_1 < m_2$ or that $\rho < \frac{1}{2}$ and $\rho > \frac{1}{2}$. Since H_0 does not specify the value of μ (which may have any value between 0 and ∞), it is a composite hypothesis in the

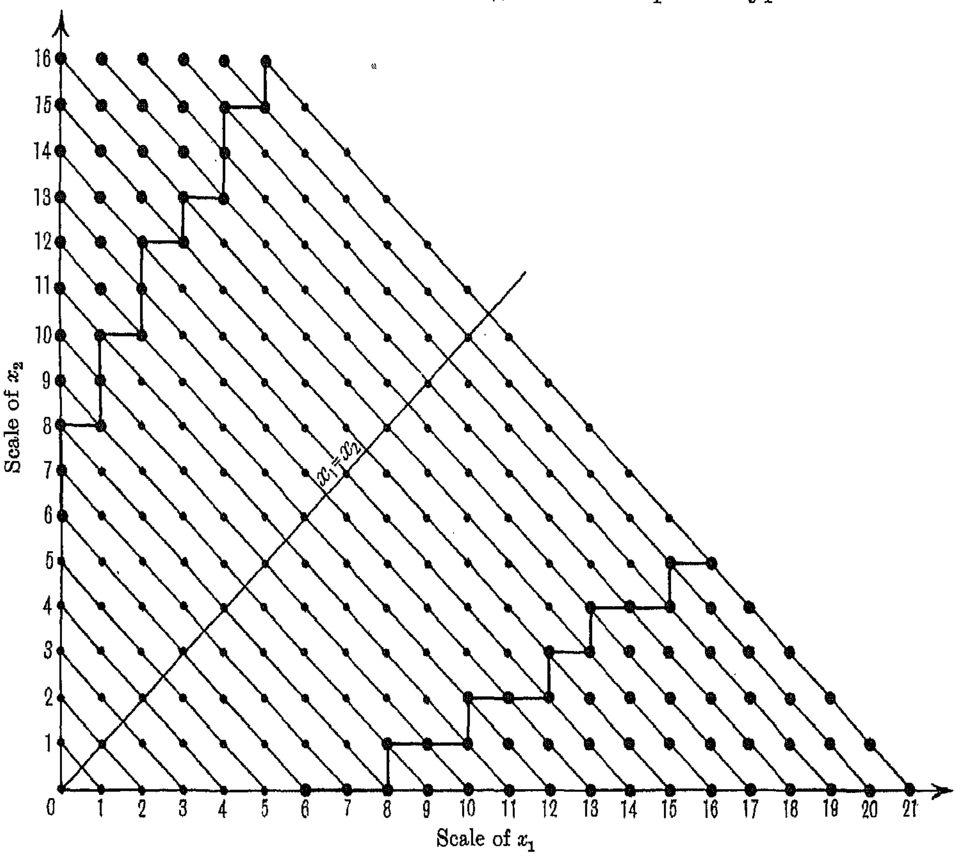


Fig. 1. Critical region for case $\alpha = 0{\cdot}05$. It consists of the two areas containing the larger dots.

sense of Neyman & Pearson. To test it, we should like to be able to use a critical region, w, such that

$$P\{E\epsilon w \mid \rho = \tfrac{1}{2}, \mu\} = \alpha, \qquad \ldots\ldots(9)$$

whatever be μ. No such region can be found, owing to the discontinuous character of the distribution of the variables, but it is possible to go a long way in determining a satisfactory region on the following lines.

It will be noted that if $\rho = \frac{1}{2}$, then the distribution of x_1 on each line, n = constant, is that of the symmetrical binomial $(\frac{1}{2}+\frac{1}{2})^n$. If ρ is either $> \frac{1}{2}$ or $< \frac{1}{2}$, we shall have a skew binomial with probability density concentrating towards the lower or upper end of the line; in order to make the test as efficient as possible in detecting departures from $\rho = \frac{1}{2}$ in either direction, the critical or rejection region should therefore include the two tails of each of the separate binomials $(\frac{1}{2}+\frac{1}{2})^n$. It may be defined as follows.

Build up the critical region w of "pieces", $w(n, \alpha)$, of the lines

$$n = x_1 + x_2 = \text{constant} \quad (n = 0, 1, 2, \ldots), \qquad \ldots\ldots(10)$$

where $w(n, \alpha)$ includes all sample points on (10) for which

$$x_1 \leq k(n, \alpha) \quad \text{and} \quad x_2 \geq n - k(n, \alpha), \qquad \ldots\ldots(11)$$

$k(n, \alpha)$ being a positive integer determined so that

(a) the sum of the terms of the binomial $(\frac{1}{2} + \frac{1}{2})^n$ corresponding to

$$0, 1, \ldots, k(n, \alpha) \text{ is } \leq \tfrac{1}{2}\alpha,$$

(b) the sum of the terms corresponding to

$$0, 1, \ldots, k(n, \alpha), k(n, \alpha) + 1 \text{ is } > \tfrac{1}{2}\alpha.$$

The pieces $w(n, \alpha)$ and part of the complete region w for the case $\alpha = 0 \cdot 05$ are indicated in Fig. 1. In view of the form of the probability law (6) it will be seen that the region w is such that

$$P\{E \epsilon w \mid p = \tfrac{1}{2}, \mu\} \leq \alpha, \qquad \ldots\ldots(12)$$

whatever be μ. Thus if we reject the hypothesis H_0 which assumes $m_1 = m_2$, whenever the sample point is included in w, i.e. falls at any of the points indicated in Fig. 1 by the larger spots, we know that the risk of the first kind of error is at most equal to α. Further, while it cannot be claimed that the test is completely unbiased or is the uniformly most powerful unbiased test,* it seems likely to be as efficient as any other alternative test in detecting departures in p from $\frac{1}{2}$.

In Table I we have given the boundary values $k(n, \alpha)$ of the critical regions associated with four values of α, viz. 0·20, 0·10, 0·05, 0·01, and for $n = x_1 + x_2$ varying from 0 to 80. Having decided on the appropriate value for α, the rule of the test is: reject the hypothesis that $m_1 = m_2$ if

$$x_1 \leq k(n, \alpha) \quad \text{or} \quad x_2 \geq n - k(n, \alpha).$$

It will be seen that for small values of n no test associated with the α required may be possible.

The power function of the test, which depends on both μ and p, may be calculated from the expression

$$P\{E \epsilon w \mid p, \mu\} = \sum_{n=0}^{\infty} \left\{ \frac{\mu^n e^{-\mu}}{n!} \sum_{w(n, \alpha)} \frac{n!}{x_1!\,(n - x_1)!} p^{x_1} (1 - p)^{n - x_1} \right\}, \qquad \ldots\ldots(13)$$

where $\sum_{w(n, \alpha)}$ is the sum of the terms of the binomial expansion

$$\{(1 - p) + p\}^n,$$

satisfying the conditions (11).

Values of the power for the two cases $\alpha = 0 \cdot 10$ and $\alpha = 0 \cdot 05$, and for a number of values of μ and p are given in Table II *a* and *b*. The computation involved (a) first finding for a given n the sums of the binomial terms in the second summation of (13) with the help of the *Tables of the Incomplete Beta Function*,

* These terms have been applied to problems of continuous variables.

(1934), and (b) then multiplying this by the appropriate term of the Poisson series, and (c) finally summing for $n = 0, 1, 2, \ldots$.

It will be noted that if the admissible alternatives to $m_1 = m_2$ are only that $m_1 > m_2$, then the appropriate critical region will be defined by $x_1 \geqslant n - k(n, \alpha)$, and the risk of the first kind of error will now be $\leqslant \frac{1}{2}\alpha$. Similarly, if the alternatives are only $m_1 < m_2$, the region will be defined by $x_1 \leqslant k(n, \alpha)$. These regions correspond to the lower and upper marked areas, respectively, of Fig. 1.

We have also given in Tables II in certain cases the actual value of $P\{E\epsilon w \mid \rho = \frac{1}{2}, \mu\}$, that is to say, the chance of rejecting the hypothesis that $m_1 = m_2$ when it is true. These chances are, as expected, considerably less than the corresponding values of α, but approach α as μ increases and the discontinuity in the distributions of x_1 and x_2 becomes of less importance. In practice we shall not, of course, know exactly what μ is, but as we shall probably have a rough idea of the size of the Poisson m in the particular problems investigated, the figures in these columns of the Tables will probably be helpful in determining the value to select for α, the upper limit of the significance level. Table III contains true significance levels for the other two limiting values, $\alpha = 0{\cdot}20$ and $\alpha = 0{\cdot}01$, for which the rejection levels $k(n, \alpha)$ have been given in Table I.

From the data of Tables II it was possible to construct the charts given in Figs. 2 and 3 (pp. 322, 323 below), showing in terms of the Poisson parameters m_1 and m_2 (rather than μ and ρ) some of the contours of equal power.

3. Discussion of tables and charts with some illustrative examples

A point which is clearly brought out by the present investigation is that it is impossible to detect differences between the means of two Poisson series m_1 and m_2, when both these quantities are small, even if one is several times larger than the other. Thus even were $m_1 = 0$ and $m_2 = 5$, the chance of detecting the difference from the samples would be only 0·384 using $\alpha = 0{\cdot}05$, and 0·560 using $\alpha = 0{\cdot}10$.* As an illustration of the kind of differences that may be detected, Fig. 2 may be used as follows. It might with reason be regarded as undesirable to plan an experiment in which the chance was less than 0·5 of detecting from two random samples that differences between m_1 and m_2 of practical importance existed. If we pass along the 50 % power contour in the diagram, we pass through points (m_1, m_2) of which the following are typical:

$$(0{\cdot}0,\ 4{\cdot}7),\quad (2{\cdot}0,\ 8{\cdot}1),\quad (6{\cdot}0,\ 14{\cdot}3),\quad (10{\cdot}0,\ 19{\cdot}8),\quad (18{\cdot}0,\ 30{\cdot}5).$$

If it is important to be able to detect differences of this magnitude, then larger samples must be taken in order to increase the expectations. If the samples

* The true values of the significance level in these cases would be 0·09 and 0·024 respectively; if the test were used with $\alpha = 0{\cdot}20$ (and a significance level in this case of 0·065) the chance of detection would be somewhat greater.

can, for instance, be increased c-fold, the expectations will be increased to cm_1, cm_2, a point lying on the same straight line passing through the origin as m_1, m_2, but placed c times further out. Since such radial lines will cut the contours of equal power at rather acute angles, considerable multiplication of sample sizes may often be needed to ensure detection of differences.

Regarding a chance of detection of 19 to 1 as satisfactory, we note that the 95 % contour in Fig. 2 passes through points (m_1, m_2) of which the following are typical:

$$(0{\cdot}0,\ 9{\cdot}1),\quad (2{\cdot}0,\ 15{\cdot}5),\quad (6{\cdot}0,\ 24{\cdot}0),\quad (10{\cdot}0,\ 31{\cdot}5).$$

Differences of this kind, if they exist, can therefore be almost certainly detected using in the test the critical values associated with $\alpha = 0{\cdot}10$. Rather smaller differences could be detected with the same high probability if α were taken as $0{\cdot}20$.

In practice, of course, expected values m_1 and m_2 are not known in advance, but we believe that reasonings of this kind based on the tables or charts may be useful as a preliminary step in planning the extent of sampling required either in experimental work or routine analysis.

Example. In the testing of clover seed for dodder it is customary to withdraw a 100 gr. sample from a sack with a long trier or probe. Clearly the number of dodder seeds found will vary from sample to sample, and from one sack to another in the same consignment. This may be due to the following causes:

(1) The material from which the samples were drawn was not thoroughly mixed.

(2) The two samples were drawn from different seed stocks.

(3) There are laboratory errors in analysis.

(4) The existence of random sampling fluctuations.

In applying statistical analysis in the comparison of two counts, the hypothesis to be tested assumes that the discrepancy to be tested is merely due to chance, and that the frequency of dodder seeds, x, will vary from one sample to another in accordance with the Poisson law. If a significant difference is found, it may of course be due to one of the first three sources of error.

Suppose that a 100 gr. sample is drawn from a sealed sack of clover and sent to a seed testing station, and that no dodder seeds are found in it, i.e. $x_1 = 0$. The buyer, before opening the sack, takes another 100 gr. sample, and on having it analysed learns that three seeds have been found, i.e. $x_2 = 3$. Would he have any justification for criticizing the first testing on the grounds that the difference between 0 and 3 could not be due to chance? On examining Table I, it is seen that for $n = x_1 + x_2 = 3$, even using the least stringent of the four tests, the difference cannot be regarded as significant.*

Suppose that on another occasion it was found that $x_1 = 2$ and $x_2 = 6$.

* It will be noted from Table III that with μ equal to 5 or less, the true significance level for the first test with $\alpha = 0{\cdot}20$ is at about $0{\cdot}06$.

Entering Table I with $n=8$, we see that for none of the levels of test does $x_1=2$ fall in the critical region. If the purchaser still considered that to find three times as many dodder seeds in one sample as in the other could not be due to chance, we might illustrate the danger of drawing conclusions in this way by a use of one of the charts. If one of the *expected* numbers were three times the other, so that $m_2=3m_1$, and therefore $\rho=0{\cdot}25$, we may ask how large $\mu=m_1+m_2$ must be before there is as much as an even chance of detecting the difference. Taking Fig. 3, it is found that the line $m_2=3m_1$ cuts the 50 % power contour at about the point $m_2=13{\cdot}8$, $m_1=4{\cdot}6$. Thus samples of clover at least twice as large as those taken would be needed to provide only a 50: 50 chance of detecting a real difference in dodder content of the order suspected by the purchaser. In this case the true significance level of the test used is seen from the number on the diagonal line of the chart to be about 0·03.

To be almost certain of detecting a difference with $m_2=3m_1$, for example to make the odds 19 to 1, we notice that a continuation of the same line cuts the 95 % power contour where $m_2=38{\cdot}4$, $m_1=12{\cdot}8$. For this samples of about 600 gr. would be needed.

4. Summary

The problem considered is that in which x_1 and x_2 are two independent random variables distributed in accordance with the Poisson law of equation (1), and it is desired to test the hypothesis that the expectations m_1 and m_2 are the same. We have shown how a test may be derived which is independent of the value of the unknown common hypothetical expectation but which, owing to the discontinuous nature of the probability distributions, will only provide an upper limit to the significance level, i.e. to the chance of rejecting the hypothesis tested when it is true. The manner of approach of the significance level to its upper limit has, however, been investigated numerically.

A table (Table I) has been provided, containing critical values $k(n, \alpha)$ required in carrying out the test. The power function of the test has also been determined, and tables (Table II *a*, *b*) and charts (Figs. 2, 3) given which make it possible to determine the chance of detecting differences in the expectations m_1 and m_2 of specified magnitudes. Finally, a discussion of some uses of the test has been added.

In conclusion we should like to express our thanks to Prof. E. S. Pearson for his helpful suggestions and criticism made during the course of our investigation.*

* [Certain modifications and additions to the paper have been made since it was received for publication at the beginning of July 1939. As circumstances have unfortunately made communication with the authors impossible, I must accept responsibility for these alterations, which have been mainly concerned with the extension of Tables I and II to higher values of n and μ. E. S. P.]

REFERENCES

NEYMAN, J. (1935). *Bull. Soc. Math. Fr.* **63**, 246.
NEYMAN, J. & PEARSON, S. E. (1933). *Philos. Trans.* A, **231**, 289.
PRZYBOROWSKI, J. & WILENSKI, H. (1935). *Biometrika*, **27**, 273.
Tables of the Incomplete Beta Function (1934). *Biometrika* publication.

TABLE I

Boundary values, $k(n, \alpha)$, for each n

$n = x_1+x_2$	Upper limit of significance level, α				$n = x_1+x_2$	Upper limit of significance level, α			
	0·20	0·10	0·05	0·01		0·20	0·10	0·05	0·01
1					41	15	14	13	11
2					42	16	15	14	12
3					43	16	15	14	12
4	0				44	17	16	15	13
5	0	0			45	17	16	15	13
6	0	0	0		46	18	16	15	13
7	1	0	0		47	18	17	16	14
8	1	1	0	0	48	19	17	16	14
9	2	1	1	0	49	19	18	17	15
10	2	1	1	0	50	19	18	17	15
11	2	2	1	0	51	20	19	18	15
12	3	2	2	1	52	20	19	18	16
13	3	3	2	1	53	21	20	18	16
14	4	3	2	1	54	21	20	19	17
15	4	3	3	2	55	22	20	19	17
16	4	4	3	2	56	22	21	20	17
17	5	4	4	2	57	23	21	20	18
18	5	5	4	3	58	23	22	21	18
19	6	5	4	3	59	24	22	21	19
20	6	5	5	3	60	24	23	21	19
21	7	6	5	4	61	24	23	22	20
22	7	6	5	4	62	25	24	22	20
23	7	7	6	4	63	25	24	23	20
24	8	7	6	5	64	26	24	23	21
25	8	7	7	5	65	26	25	24	21
26	9	8	7	6	66	27	25	24	22
27	9	8	7	6	67	27	26	25	22
28	10	9	8	6	68	28	26	25	22
29	10	9	8	7	69	28	27	25	23
30	10	10	9	7	70	29	27	26	23
31	11	10	9	7	71	29	28	26	24
32	11	10	9	8	72	30	28	27	24
33	12	11	10	8	73	30	28	27	25
34	12	11	10	9	74	30	29	28	25
35	13	12	11	9	75	31	29	28	25
36	13	12	11	9	76	31	30	28	26
37	14	13	12	10	77	32	30	29	26
38	14	13	12	10	78	32	31	29	27
39	15	13	12	11	79	33	31	30	27
40	15	14	13	11	80	33	32	30	28

Rule of test: reject hypothesis that $m_1 = m_2$ if $x_1 \leqslant k(n, \alpha)$ or $x_1 \geqslant n - k(n, \alpha)$.

See note on p. 323 regarding extension of limits for $n > 80$.

TABLE III

Upper limit of significance level $\alpha = 0{\cdot}20$		Upper limit of significance level $\alpha = 0{\cdot}01$	
μ	True level	μ	True level
5	0·065	5	0·0007
10	0·109	10	0·0031
15	0·123	15	0·0042
20	0·132	20	0·0048
25	0·137	25	0·0054
30	0·141	30	0·0058
35	0·148	35	0·0062
40	0·152	40	0·0065
45	0·154	45	0·0067
50	0·155	50	0·0068

N.B. If hypothesis tested is true, $\mu = 2m_1 = 2m_2$.

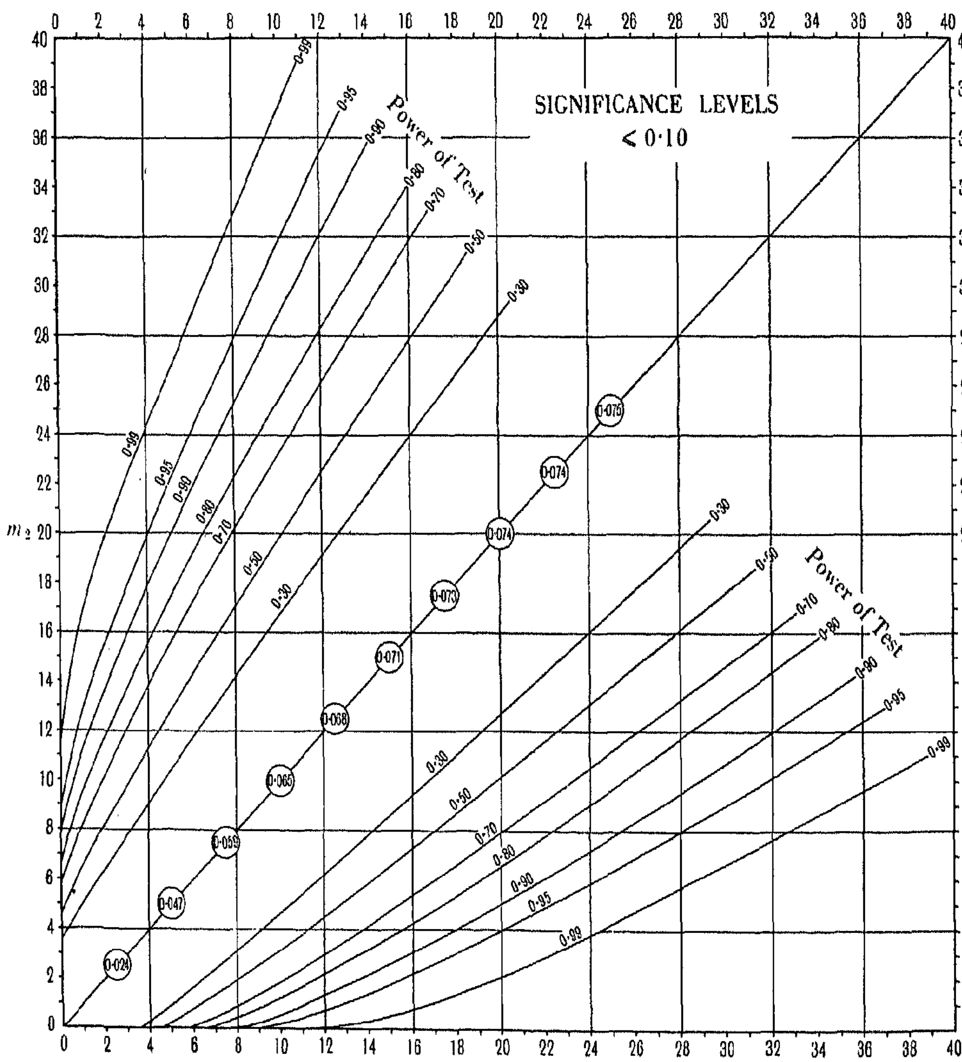


Fig. 2

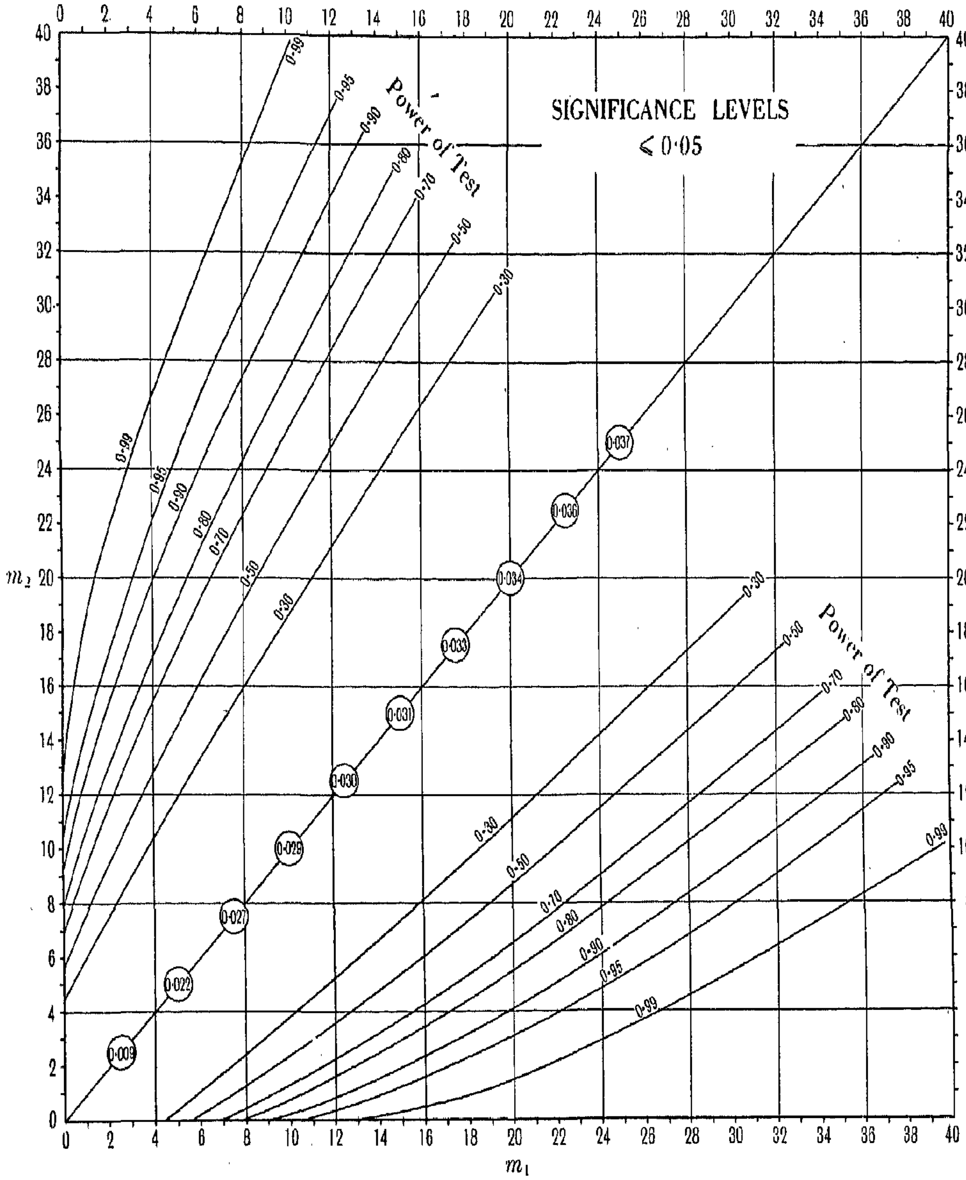


Fig. 3

Note regarding Table I. When x_1 and x_2 are large, if the hypothesis tested is true then $(x_1-x_2)/\sqrt{(x_1+x_2)}$ may be regarded as a unit normal deviate or, from another point of view, x_1 should be normally distributed about $\frac{1}{2}n$ with standard deviation $\frac{1}{2}\sqrt{n}$. On this basis, in the case of $n=80$, the values of $k(n,\alpha)$ for $\alpha=0{\cdot}20$, $0{\cdot}10$, $0{\cdot}05$ and $0{\cdot}01$ are found to be 34·3, 32·6, 31·2 and 28·5, respectively. Applying a correction for discontinuity we have 33, 32, 30 and 28 for the integral values of $k(n,\alpha)$ associated with upper limits of significance level equal to the four values of α. These are the values for $k(n,\alpha)$ given in the last row of Table I. It is clear that beyond $n=80$ the normal approximation is valid.

ON THE METHOD OF PAIRED COMPARISONS

By M. G. KENDALL and B. BABINGTON SMITH

Introduction

1. Suppose we have a number of objects A, B, C, etc. which are to be considered according to the different degrees in which they exhibit some common quality. If the quality is measurable in some objective way the objects will yield a number of variate values, in which case the problem is amenable to treatment by well-understood methods. It may, however, happen either for theoretical or for practical reasons that the quality is not measurable. We then have to rely for a discussion of the variation of the quality on judgments of a more or less subjective kind carried out after a comparison of the objects among themselves.

One of the methods of comparison which has been widely used in this connexion is that of ranking. An observer examines the objects and arranges them in the order in which he judges them to possess the quality under consideration. This arrangement is called a ranking, and when two or more observers provide rankings of the same set of objects there arise the familiar questions of the type: is there any significant resemblance between the judgments of observers? or, do the data furnish any evidence that the objects have a "real" objective ranking?

2. The ranking method suffers from a serious drawback when the quality considered is not known with certainty to be representable by a linear variable. We may, for instance, ask an observer to rank a number of individuals in order of intelligence, and he may comply with the request in the full belief that he is doing something within his powers; but if intelligence is not measurable on a linear scale this ranking may fail to give a real picture either of the observer's preference or of the variation of intelligence among the individuals. It is not impossible that the observer should judge A more intelligent than B, B than C, and C than A, if the individuals are presented for his consideration one pair at a time. The likelihood of this happening is obviously increased when we are dealing with tastes in music, eatables or film stars; and in practice the event is not uncommon. Such "inconsistent" preferences can never appear in ranking, for if A is preferred to B and B to C, then A must automatically be shown as preferred to C. The use of ranking thus destroys what may be valuable information about preferences.

3. In this paper we consider a more general method of investigating preferences. With n objects, we shall suppose that each of the $\binom{n}{2}$ possible pairs is presented to an observer and his preference of one member of the pair noted.

We assume that a choice between two objects can always be made.* With m observers the data then comprise $m\binom{n}{2}$ preferences. The questions to be discussed include:

(a) Is there any evidence that a particular observer is capable of forming a reliable judgment of the quality under investigation; and if not, is the fault his, or is it due to the fact that he has been asked to perform an impossible task?

(b) Is there any significant concordance of preferences between observers?

(c) Can the quality under discussion be represented by a linear variable?

4. The method of offering for judgment objects two at a time is known as the method of paired comparisons. Hitherto it has been used mainly in human psychology, but it has some interesting applications in animal experimentation. For instance, in feeding experiments it is impossible to get an animal to rank a number of foods in order of preference but it is not difficult to offer pairs of foods and to note which is taken first. Experiments of this kind have, of course, to be conducted with great care to ensure that conditions operating when the different pairs are offered are as constant as possible; but the difficulties are far from being insuperable and the method of paired comparisons offers a useful technique in cases where the more usual procedures cannot be applied. From the point of view of theoretical statistics perhaps the most interesting part of the present work is that it offers some lines of approach to the difficult question whether a given quality can be legitimately regarded as based on a linear variable, i.e. whether ranking or scoring methods are justifiable or not.

Consistence in preferences

5. If the object A is preferred to B we write $A \rightarrow B$ or $B \leftarrow A$. The $\binom{n}{2}$ preferences of a single observer may be represented in tabular form as shown in Table I.

In this table, which is shown for the six objects A to F, an entry of unity in column Y and row X means $X \rightarrow Y$, and is thus accompanied by a complementary zero in row Y and column X. The diagonals are blocked out. For example, in Table I, $A \rightarrow B$, $A \rightarrow C$, $D \rightarrow A$, etc.

The arrangement of the objects A to F in the row and column headings is quite arbitrary. There are $(n!)^2$ ways of representing the same configuration of preferences in such a table according to the permutations of objects in row and

* That is, we exclude cases in which an observer cannot make up his mind which object he prefers, just as in the ranking case one excludes the possibility of split ranks. In practice it sometimes happens that an observer is genuinely unable to reach a decision. To allow for this fact in the theoretical discussions would introduce complications of a most intractable kind. When the effect becomes important in practice it can be allowed for by selecting the set of preferences which are most unfavourable to the hypothesis under test.

TABLE I

	A	B	C	D	E	F
A	—	1	1	0	1	1
B	0	—	0	1	1	0
C	0	1	—	1	1	1
D	1	0	0	—	0	0
E	0	0	0	1	—	1
F	0	1	0	1	0	—

column; but in practice it is generally desirable to have the order in row and column the same, and even among the $n!$ possible arrangements so given there are often practical considerations which determine one order as more convenient than others.

6. Paired comparisons may also be represented geometrically by a method which can be illustrated for the case of the six objects as follows:

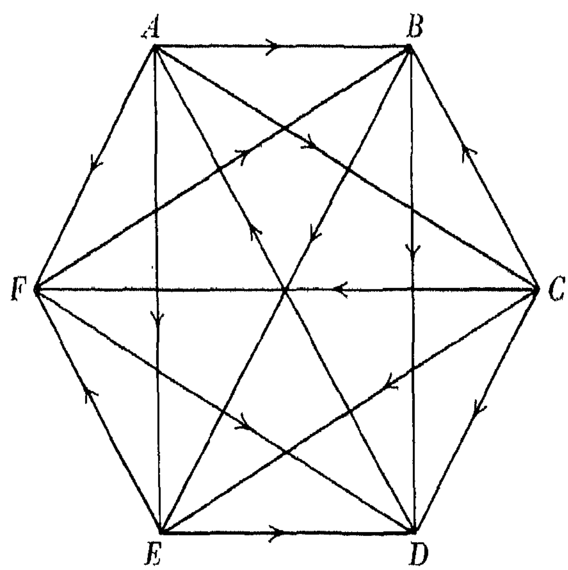


Fig. 1. Geometrical representations of the scheme of preferences of Table I.

We represent the six objects A to F by the six vertices of a regular hexagon and join the vertices in all possible ways by straight lines. If $A \rightarrow B$ we draw

an arrow on the line AB pointing from A to B. The arrows shown on Fig. 1 correspond to the preferences shown in Table I.*

7. If an observer makes preferences of type $A \rightarrow B \rightarrow C \rightarrow A$ we say that the triad ABC is inconsistent. In the geometrical representation an inconsistent triad is shown by a triangle in which all the arrows go round in the same direction. We may thus speak of a "circular" triad of preferences. In Fig. 1 the triads ACD, BEF and three others are circular.

It is also possible to have inconsistent triads of greater extent; but any such circuit must contain at least two circular triads. Suppose, for instance, that $ABCD$ is circular, e.g. that $A \rightarrow B \rightarrow C \rightarrow D \rightarrow A$. Then either $A \rightarrow C$ or $C \rightarrow A$. In the first case ACD is circular, in the second ABC. Similarly either ABD or BCD is circular. Thus the circular tetrad must contain just two circular triads. On the other hand it is possible for a tetrad to contain circular triads without being itself circular.

Similarly, if $ABCDE$ is circular either ABC or $ACDE$ is circular and either BCD or $BDEA$ is circular. If the two tetrads are circular there must be at least three circular triads (not necessarily four, because ADE may be common to both). It is easy to see by an actual example based on this configuration that there need not be more than three circular triads; and it is clear that there must be at least three. For if the tetrads are not circular then ABC and BCD must be so and then either CDE is circular or $ABCE$ is so, adding at least one more.

Generally, it appears that a circular n-ad must contain at least $(n-2)$ circular triads; but it may contain more, and the fact that an n-ad contains $(n-2)$ circular triads does not mean that it is itself circular. In discussing inconsistences, therefore, it seems best to confine attention to circular triads, which, so to speak, constitute the inconsistent elements of the configuration, and to ignore the more ambiguous criteria associated with circular polyads of greater extent.

8. We now prove the following theorems:

(1) The maximum possible number of circular triads is $(n^3-n)/24$ if n is odd and $(n^3-4n)/24$ if n is even; and the minimum number is zero.

(2) These limits can always be attained by some configuration of preferences.

(3) For any integral number between the maximum and the minimum there exists at least one preference-configuration with that number of circular triads; and in general there will be more than one.

Consider a polygon of the type shown in Fig. 1 with n vertices. There will

* These preferences were obtained in an experiment on a dog, which was offered the following foods in pairs: meat, biscuit, chocolate, apple, pear and cheese. The members of a pair were cut to the same size and placed equidistantly from the dog, which was then released and allowed to choose. All the pieces of food were eaten avidly, it being that sort of dog, but there were considerable inconsistences in choice. We do not offer these data as more than an illustration of the method.

be $(n-1)$ lines emanating from each vertex. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the number of lines at the respective vertices on which the arrows *leave* the vertex.

Then
$$\mathop{S}_{r=1}^{n}(\alpha_r) = \binom{n}{2}$$
and the mean value of α_r is $(n-1)/2$.

Define
$$T = \mathop{S}_{r=1}^{n}\left(\alpha_r - \frac{n-1}{2}\right)^2$$
$$= S(\alpha_r^2) - \frac{n(n-1)^2}{4}. \qquad \ldots\ldots(1)$$

We now show that if the direction of a preference is altered and the effect is to increase the number of circular triads by d, T is reduced by $2d$; and conversely. Consider the preference $A\rightarrow B$. The only triads affected by altering this to $B\rightarrow A$ are those containing the line AB. Suppose there are α preferences of type $A\rightarrow X$ (including $A\rightarrow B$) and β preferences of type $B\rightarrow X$. Then four possible types of triad arise:

$A\rightarrow X\leftarrow B$, say p in number
$A\leftarrow X\rightarrow B$,
$A\rightarrow X\rightarrow B$, which must number $\alpha-p-1$
$A\leftarrow X\leftarrow B$, ,, ,, $\beta-p$.

When the preference $A\rightarrow B$ is reversed the first two remain non-circular. The third becomes circular, the fourth ceases to be so. The reduction in the value of T is
$$\alpha^2-(\alpha-1)^2+\beta^2-(\beta+1)^2$$
$$=2(\alpha-\beta-1)$$
$$=2d, \text{ say.}$$

The increase in the number of circular triads is
$$(\alpha-p-1)-(\beta-p)=\alpha-\beta-1$$
$$=d.$$

More generally, if as the result of reversing any number of preferences T is decreased by $2d$, then d must be an integer and the number of circular triads must be increased by d. This clearly follows from the previous results for the reversal of preferences can take place one at a time and the effect on T and the number of circular triads is cumulative.

We now investigate the maximum and minimum values of T. It is clear from the definition that T is greatest when the α's are the natural numbers $1, 2, \ldots, n$; and this is a possible case because it corresponds to ordinary ranking. Hence max. $(T) = (n^3-n)/12$.

For the minimum value, consider the polygon $A_1, A_2, \ldots, A_n$. Set up the preferences $A_1\rightarrow A_2\rightarrow \ldots A_n\rightarrow A_1$. Clearly at any vertex this results in one arrow

entering and one leaving the vertex, i.e. the contribution to α is unity at each vertex. Next set up the preferences $A_1 \to A_3 \to A_5 \to \ldots$. This circuit may either visit each vertex once, or not. In the latter case we proceed to an unvisited vertex and set up the preferences $A_r \to A_{r+2} \to A_{r+4} \to \ldots$ and so on. Again there will be a unit contribution to all the α's.

We then set up the preferences $A_1 \to A_4 \to A_7 \to$ etc. and so on; and in this way we shall ultimately complete the preference scheme.

If n is odd all the preferences described will consist of circular tours of the polygon, and thus the value of α for each vertex will be $(n-1)/2$. If n is even the last preference $A_1 \to A_{\frac{1}{2}n+1}$ will not be a tour but will consist of the single line joining one vertex with the symmetrically opposite vertex. Thus there will be $n/2$ vertices for which $\alpha = n/2$ and $n/2$ vertices for which $\alpha = (n-2)/2$. In this case $T = n/4$.

Now it is clear from the definition of T that it cannot be less than zero, or if n is even, be less than $n/4$. The configuration just given shows that these minima are, in fact, attainable.

Thus T can vary from a maximum of $(n^3-n)/12$ to a minimum of zero or $n/4$. Hence the maximum number of circular triads, being half the variation from maximum to minimum of T (the maximum of T corresponding to the ranking case in which there are no inconsistences) is $(n^3-4n)/24$ if n is even and $(n^3-n)/24$ if n is odd.

This establishes the first two results enunciated at the beginning of this section. To prove the third it is sufficient to give a systematic method of proceeding from the configuration of minimum to that of maximum inconsistence by steps decreasing T two at a time. Consider, for example, the case $n = 8$. For the minimum inconsistence the α's will have the values 0 to 7, which we set out thus:

A	B	C	D	E	F	G	H
0	1	2	3	4	5	6	7

We proceed by reversing the preferences between vertices whose α-values differ by two. This clearly reduces T by two.

Reversing the preferences between C and E we get

A	B	C	D	E	F	G	H
0	1	3	3	3	5	6	7

and between D and F we get

A	B	C	D	E	F	G	H
0	1	3	4	3	4	6	7

which we may rearrange as

A	B	C	E	D	F	G	H
0	1	3	3	4	4	6	7

Now reversing the preferences between B and E and between D and G and rearranging we have

A	B	E	C	F	D	G	H
0	2	2	3	4	5	5	7

and now interchanging A and B, G and H,

A	B	E	C	F	D	G	H
1	1	2	3	4	5	6	6

At this stage we have preserved the α-numbers 2, 3, 4 and 5 in the middle but reduced the extremes A and H. We can now carry out the process again, arriving at the α-numbers.

$$1 \quad 2 \quad 2 \quad 3 \quad 4 \quad 5 \quad 5 \quad 6$$

and twice again, giving

$$2 \quad 2 \quad 3 \quad 3 \quad 4 \quad 4 \quad 5 \quad 5$$

whence a final interchange gives

$$3 \quad 3 \quad 3 \quad 3 \quad 4 \quad 4 \quad 4 \quad 4$$

and this is the position of maximum inconsistence. It is readily verified by following the interchanges on a polygon diagram that the reversals are, in fact, legitimate.

Coefficient of consistence in paired comparisons

9. If d is the number of circular triads in an observed configuration of preferences we define

$$\left.\begin{aligned} \zeta &= 1-\frac{24d}{n^3-n}, \quad n \text{ odd} \\ &= 1-\frac{24d}{n^3-4n}, \quad n \text{ even} \end{aligned}\right\} \qquad \ldots\ldots(2)$$

and call ζ the coefficient of consistence. If and only if it is unity there are no inconsistences in the configuration, which may therefore be represented by a ranking. As ζ decreases to zero the inconsistence, as measured by the number of circular triads, increases.

For example, in the configuration of Fig. 1 there are five circular triads, *ABD*, *ACD*, *AFD*, *AED* and *BEF*. The maximum possible number is 8. Thus $\zeta = 0{\cdot}375$.

10. ζ can also be interpreted in the light of Table I. Suppose, in that table, we sum the rows. (The column sums are determined by the row sums and add no fresh information.) The sum of any row will be the α-number for that vertex in the polygon which corresponds to the object defining the row. T will then be the value of the sum of squares of deviations of row totals from the mean value $(n-1)/2$, that is to say, will be the variance of the row sums multiplied

by n. ζ is thus a linear function of this variance; but it cannot be tested in the χ^2 distribution as if Table I were a contingency table, for the border cells are not independent or linearly dependent.

11. If an individual observer produces a configuration of preferences which show inconsistence there are usually several explanations; he may be an incompetent judge, the objects may be so alike that consistent differentiation is not possible, or his attention may wander during the course of the experiment. We discuss these questions later. They are mentioned here to explain the motive for the next stage of the mathematics. With what probability can a value of ζ arise by chance if the observer allots his preferences at random with respect to the quality under consideration?

With n objects there are $2^{\binom{n}{2}}$ possible configurations of preferences. We proceed to investigate the distribution of d in this universe of $2^{\binom{n}{2}}$ different members. The method consists of proceeding from the distribution for n to that for $(n+1)$.

For $n=3$ there are eight configurations, of which two give one circular triad and six no circular triads. Consider the effect of adding a new vertex D to the vertices ABC. Four cases arise:

(1) $D\rightarrow$ all A, B, C.
(2) $D\rightarrow$ two of A, B, C.
(3) $D\rightarrow$ one of A, B, C.
(4) $D\rightarrow$ none of A, B, C.

The last two are symmetrical with the first two and need not be separately considered.

Situation (1) arises in one way and clearly does not add any new circular triads other than those already existing in the configuration ABC. It therefore contributes six values $d=0$ and two values $d=1$. So does situation (4).

Situation (2) arises in three ways, according as $D\leftarrow A$, B, or C. The configurations so reached are similar and we may take any one, say $D\leftarrow C$, as the single preference. If $A\leftarrow C$ then DAC is not circular and if $B\leftarrow C$ the DBC is not circular. On the other hand $A\rightarrow C$ and $B\rightarrow C$ will each produce a circular triad. We then have the cases

	No. of circular triplets added
$A\leftarrow C\rightarrow B$	0
$A\rightarrow C\rightarrow B$	1
$A\leftarrow C\leftarrow B$	1
$A\rightarrow C\leftarrow B$	2

We now consider AB. In the first two cases just enumerated the direction of AB does not matter and no circular triads are added. With the third $A \rightarrow B$ gives no circular triad but $A \leftarrow B$ adds one. With the fourth $A \rightarrow B$ adds one and $A \leftarrow B$ adds none.

Thus the number of circular triads occurring for these four cases is found to be

No. of circular triplets	Frequency
0	2
1	2
2	4

We must multiply the frequency by three and by two to allow for similar symmetrical arrangements, and the final results are

No. of circular triplets	Frequency
0	24
1	16
2	24
Total	64

The principles of this method are clear enough and the work may be formalized by a number of conventions which we omit to save space. In common with many similar combinatorial problems, however, troubles arise from the sheer number of possibilities and the difficulty of ensuring that nothing is overlooked. Up to the present we have found the distribution of d for n up to and including 7. The frequencies and probabilities are given in Table II.

12. For the values already obtained the moments are given by the following formulae:

$$\mu_1' \text{ (about 0)} = \frac{1}{4}\binom{n}{3}, \qquad \text{......(3)}$$

$$\mu_2 = \frac{3}{16}\binom{n}{3}, \qquad \text{......(4)}$$

$$\mu_3 = -\frac{3}{32}\binom{n}{3}(n-4), \qquad \text{......(5)}$$

$$\mu_4 = \frac{3}{256}\binom{n}{3}\left\{9\binom{n-3}{3}+39\binom{n-3}{2}+9\binom{n-3}{1}+7\right\}. \qquad \text{......(6)}$$

We have very little doubt that these results are true in general but can offer no rigorous proof. In so far, however, as the moments are in a sense symmetric sums it appears highly probable that they are given by polynomials

TABLE II

Frequency (f) of values of d and probability (P) that values will be attained or exceeded

Value of d	$n=2$		$n=3$		$n=4$		$n=5$		$n=6$		$n=7$	
	f	P	f	P	f	P	f	P	f	P	f	P
0	2	1·000	6	1·000	24	1·000	120	1·000	720	1·000	5,040	1·000
1			2	0·250	16	·625	120	·883	960	·978	8,400	·998
2					24	·375	240	·766	2,240	·949	21,840	·994
3							240	·531	2,880	·880	33,600	·983
4							280	·297	6,240	·792	75,600	·967
5							24	·023	3,648	·602	90,384	·931
6									8,640	·491	179,760	·888
7									4,800	·227	188,160	·802
8									2,640	·081	277,200	·713
9											280,560	·580
10											384,048	·447
11											244,160	·263
12											233,520	·147
13											72,240	·036
14											2,640	·001
Total	2	—	8	—	64	—	1,024	—	32,768	—	2,097,152	—

in n; and if this is so the values obtained are sufficient to establish polynomials of degree six or less.

It is also to be noted that from the above values of the moments

$$\beta_1 = \mu_3^2/\mu_2^3 \sim 8/n, \quad \beta_2 = \mu_4/\mu_2^2 \sim 3 + 12/n,$$

from which it appears that a Type III distribution would fit the d-distribution fairly closely for moderate or large values of n. But as the distribution of d is of interest mainly for low values of n, which are all that occur in practice, it hardly seems worth while attempting to fit a curve.

Agreement among several observers

13. We now consider the investigation of similarities of judgments for m observers. Suppose that in a table of the form of Table I we enter a unit in the cell in row X and column Y whenever $X \rightarrow Y$ and count the units in each cell. A cell may then contain any number from 0 to m. If the observers are in complete agreement there will be $\binom{n}{2}$ cells containing the number m, the remaining $\binom{n}{2}$ cells being zero. The agreement may be complete even if there are inconsistences present.

Suppose that the cell in row X and column Y contains the number γ. Let

$$\Sigma = S\binom{\gamma}{2}, \qquad \text{......(7)}$$

the summation extending over the $n(n-1)$ cells of the table (the diagonal cells being ignored). Σ is then the sum of the number of agreements between pairs of judges. Put

$$u = \frac{2\Sigma}{\binom{m}{2}\binom{n}{2}} - 1. \qquad \text{......(8)}$$

The maximum number of agreements, occurring if $\binom{n}{2}$ cells each contain m, is $\binom{n}{2}\binom{m}{2}$ and thus in the case of complete agreement, and only in this case, $u = 1$. The further we go from this case, as measured by agreements between pairs of observers, the smaller u becomes. The minimum number of agreements occurs when each cell contains $m/2$ if m is even or $(m \pm 1)/2$ if m is odd. That is, if m is even, the minimum number of agreements is

$$2\binom{\frac{m}{2}}{2}\binom{n}{2} = \tfrac{1}{4}m(m-2)\binom{n}{2},$$

and in this case

$$u = -\frac{1}{m-1}. \qquad \text{......(9)}$$

When m is odd the minimum value of u is found to be

$$u = -\frac{1}{m}. \qquad \text{......(10)}$$

14. We propose to call u the coefficient of agreement. It is unity if and only if there is complete agreement in the comparisons. Its minimum value is not -1 except when $m = 2$. This, however, is to be expected in a measure of agreement for there can be no such thing as complete disagreement among three or more observers in paired comparisons. If observer P differs in certain comparisons from observers Q and R, the two latter must agree on those comparisons.

When $m = 2$, u reduces to

$$u = \frac{2\Sigma}{\binom{n}{2}} - 1 \qquad \text{......(11)}$$

and Σ becomes twice the number of cases in which the two observers agree about a comparison. u is thus a generalization of a coefficient τ proposed by Kendall (1938) to measure the correlation between two rankings. For general m, if the entries in the table were constrained to the ranking type, u would be the average intercorrelation τ between observers taken two at a time.

15. In discussing the significance of u it is desirable to know whether the set of preferences which give rise to it could have arisen by chance if the preferences had been assigned at random with respect to the quality under consideration. The procedure which first suggests itself is a generalization of the method used for the case of m rankings (Kendall & Babington Smith, 1939). That is to say, we sum the entries in the rows of the table and consider the variance of these entries. If the preferences are allotted at random we expect to find about equal numbers given to each object, and the variance will be low; in other cases it will be higher.

The difficulty about this suggestion is that it has not been found possible to ascertain the distribution of the variance in the $2^{m\binom{n}{2}}$ possible sets of preferences. The case $m = 1$, corresponding to the distribution of d for inconsistences, is difficult enough to solve. For higher values of m we have failed to find any distributions except in trivial cases.

We can, however, offer a test based on the distribution of u (or Σ). The comparative simplicity of the distributions in this case is in accordance with the remark made by Kendall in the paper under reference that the distribution of τ is much simpler and much more regular than the distribution of the Spearman correlation coefficient ρ.

16. Consider one cell in the table in row X and column Y and let it contain the number γ. Then the corresponding cell in row Y and column X will contain $m-\gamma$. Thus these two contribute to Σ the amount $\binom{\gamma}{2}+\binom{m-\gamma}{2}$.

Now, of the total ways in which the units can be distributed in the first cell there will be $\binom{m}{\gamma}$ in which γ units occur. Consequently the distribution of Σ in the cell and the corresponding cell is given by the expression

$$f = t^{\binom{m}{2}} + \binom{m}{1} t^{\binom{m-1}{2}} + \binom{m}{2} t^{\binom{m-2}{2}+\binom{2}{2}} + \dots + \binom{m}{\gamma} t^{\binom{m-\gamma}{2}+\binom{\gamma}{2}} + \dots + t^{\binom{m}{2}}, \quad \text{......(12)}$$

and since the distribution in other pairs of cells is independent if the preferences are allotted at random the distribution of Σ for the whole table is given by

$$D(\Sigma) = f^N, \quad \text{......(13)}$$

where $N = \binom{n}{2}$.

17. The distributions have been worked out for the following values of m and n: $m = 3$, $n = 2$ to 8; $m = 4$, $n = 2$ to 6; $m = 5$, $n = 2$ to 5; $m = 6$, $n = 2$ to 4. Tables III to VI give the probabilities based on these distributions, i.e. the probabilities that a given value of Σ will be attained or exceeded.

For constant n the distribution tends to the Type III form as m tends to infinity. In fact, for a single pair of related cells the variate value corresponding

TABLE III

The probability P that a value of Σ will be attained or exceeded, for $m = 3$, $n = 2$ to 8

$n=2$		$n=3$		$n=4$		$n=5$		$n=6$		$n=7$		$n=8$	
Σ	P	Σ	P	Σ	P	Σ	P	Σ	P	Σ	P	Σ	P
1	1·000	3	1·000	6	1·000	10	1·000	15	1·000	21	1·000	28	1·000
3	·250	5	·578	8	·822	12	·944	17	·987	23	·998	30	1·000
		7	·156	10	·466	14	·756	19	·920	25	·981	32	·997
		9	·016	12	·169	16	·474	21	·764	27	·925	34	·983
				14	·038	18	·224	23	·539	29	·808	36	·945
				16	·0046	20	·078	25	·314	31	·633	38	·865
				18	$\cdot 0^{3}24$	22	·020	27	·148	33	·433	40	·736
						24	·0035	29	·057	35	·256	42	·572
						26	$\cdot 0^{3}42$	31	·017	37	·130	44	·400
						28	$\cdot 0^{4}30$	33	·0042	39	·056	46	·250
						30	$\cdot 0^{6}95$	35	$\cdot 0^{3}79$	41	·021	48	·138
								37	$\cdot 0^{3}12$	43	·0064	50	·068
								39	$\cdot 0^{4}12$	45	·0017	52	·029
								41	$\cdot 0^{6}92$	47	$\cdot 0^{3}37$	54	·011
								43	$\cdot 0^{7}43$	49	$\cdot 0^{4}68$	56	·0038
								45	$\cdot 0^{9}93$	51	$\cdot 0^{4}10$	58	·0011
										53	$\cdot 0^{5}12$	60	$\cdot 0^{3}29$
										55	$\cdot 0^{6}12$	62	$\cdot 0^{4}66$
										57	$\cdot 0^{8}86$	64	$\cdot 0^{4}13$
										59	$\cdot 0^{9}44$	66	$\cdot 0^{5}22$
										61	$\cdot 0^{10}15$	68	$\cdot 0^{6}32$
										63	$\cdot 0^{12}23$	70	$\cdot 0^{7}40$
												72	$\cdot 0^{8}42$
												74	$\cdot 0^{9}36$
												76	$\cdot 0^{10}24$
												78	$\cdot 0^{11}13$
												80	$\cdot 0^{13}48$
												82	$\cdot 0^{14}12$
												84	$\cdot 0^{16}14$

TABLE IV

The probability P that a value of Σ will be attained or exceeded, for $m = 4$ and $n = 2$ to 6 (for $n = 6$ only values beyond the 1 % point are given)

$n=2$		$n=3$		$n=4$		$n=5$		$n=5$		$n=6$		$n=6$	
Σ	P	Σ	P	Σ	P	Σ	P	Σ	P	Σ	P	Σ	P
2	1·000	6	1·000	12	1·000	20	1·000	42	·0048	57	·014	79	$\cdot 0^{8}42$
3	·625	7	·947	13	·997	21	1·000	43	·0030	58	·0092	80	$\cdot 0^{8}28$
6	·125	8	·736	14	·975	22	·999	44	·0017	59	·0058	81	$\cdot 0^{9}98$
		9	·455	15	·901	23	·995	45	$\cdot 0^{3}73$	60	·0037	82	$\cdot 0^{9}15$
		10	·330	16	·769	24	·979	46	$\cdot 0^{3}41$	61	·0022	83	$\cdot 0^{9}12$
		11	·277	17	·632	25	·942	47	$\cdot 0^{3}24$	62	·0013	84	$\cdot 0^{10}51$
		12	·137	18	·524	26	·882	48	$\cdot 0^{4}90$	63	$\cdot 0^{3}76$	86	$\cdot 0^{11}30$
		14	·043	19	·410	27	·805	49	$\cdot 0^{4}37$	64	$\cdot 0^{3}44$	87	$\cdot 0^{11}17$
		15	·025	20	·278	28	·719	50	$\cdot 0^{4}25$	65	$\cdot 0^{3}23$	90	$\cdot 0^{13}28$
		18	·0020	21	·185	29	·621	51	$\cdot 0^{5}93$	66	$\cdot 0^{3}13$		
				22	·137	30	·514	52	$\cdot 0^{5}21$	67	$\cdot 0^{4}72$		
				23	·088	31	·413	53	$\cdot 0^{5}17$	68	$\cdot 0^{4}36$		
				24	·044	32	·327	54	$\cdot 0^{6}74$	69	$\cdot 0^{4}18$		
				25	·027	33	·249	56	$\cdot 0^{7}66$	70	$\cdot 0^{5}97$		
				26	·019	34	·179	57	$\cdot 0^{7}38$	71	$\cdot 0^{5}47$		
				27	·0079	35	·127	60	$\cdot 0^{9}93$	72	$\cdot 0^{5}20$		
				28	·0030	36	·090			73	$\cdot 0^{5}10$		
				29	·0025	37	·060			74	$\cdot 0^{6}51$		
				30	·0011	38	·038			75	$\cdot 0^{6}18$		
				32	$\cdot 0^{3}16$	39	·024			76	$\cdot 0^{7}78$		
				33	$\cdot 0^{4}95$	40	·016			77	$\cdot 0^{7}44$		
				36	$\cdot 0^{5}38$	41	·0088			78	$\cdot 0^{7}15$		

TABLE V

The probability P that a value of Σ will be attained or exceeded, for $m = 5$ and $n = 2$ to 5

$n=2$		$n=3$		$n=4$		$n=5$		$n=5$	
Σ	P	Σ	P	Σ	P	Σ	P	Σ	P
4	1·000	12	1·000	24	1·000	40	1·000	76	$\cdot 0^{4}50$
6	·375	14	·756	26	·940	42	·991	78	$\cdot 0^{4}16$
10	·063	16	·390	28	·762	44	·945	80	$\cdot 0^{5}50$
		18	·207	30	·538	46	·843	82	$\cdot 0^{5}15$
		20	·103	32	·353	48	·698	84	$\cdot 0^{6}39$
		22	·030	34	·208	50	·537	86	$\cdot 0^{6}10$
		24	·011	36	·107	52	·384	88	$\cdot 0^{7}23$
		26	·0039	38	·053	54	·254	90	$\cdot 0^{8}53$
		30	$\cdot 0^{3}24$	40	·024	56	·158	92	$\cdot 0^{8}12$
				42	·0093	58	·092	94	$\cdot 0^{9}14$
				44	·0036	60	·050	96	$\cdot 0^{10}46$
				46	·0012	62	·026	100	$\cdot 0^{12}91$
				48	$\cdot 0^{3}36$	64	·012		
				50	$\cdot 0^{3}12$	66	·0057		
				52	$\cdot 0^{4}28$	68	·0025		
				54	$\cdot 0^{5}54$	70	·0010		
				56	$\cdot 0^{5}18$	72	$\cdot 0^{3}39$		
				60	$\cdot 0^{7}60$	74	$\cdot 0^{3}14$		

TABLE VI

The probability P that a value of Σ will be attained or exceeded, for $m = 6$ and $n = 2$ to 4

$n=2$		$n=3$		$n=4$		$n=4$		$n=4$	
Σ	P	Σ	P	Σ	P	Σ	P	Σ	P
6	1·000	18	1·000	36	1·000	55	·043	74	$\cdot 0^{4}12$
7	·688	19	·969	37	·999	56	·029	75	$\cdot 0^{5}89$
10	·219	20	·832	38	·991	57	·020	76	$\cdot 0^{5}49$
15	·031	21	·626	39	·959	58	·016	77	$\cdot 0^{5}32$
		22	·523	40	·896	59	·011	80	$\cdot 0^{6}68$
		23	·468	41	·822	60	·0072	81	$\cdot 0^{6}17$
		24	·303	42	·755	61	·0049	82	$\cdot 0^{6}12$
		26	·180	43	·669	62	·0034	85	$\cdot 0^{7}34$
		27	·147	44	·556	63	·0025	90	$\cdot 0^{8}93$
		28	·088	45	·466	64	·0016		
		29	·061	46	·409	65	$\cdot 0^{3}83$		
		30	·040	47	·337	66	$\cdot 0^{3}66$		
		31	·034	48	·257	67	$\cdot 0^{3}48$		
		32	·023	49	·209	68	$\cdot 0^{3}26$		
		35	·0062	50	·175	69	$\cdot 0^{3}16$		
		36	·0029	51	·133	70	$\cdot 0^{4}86$		
		37	·0020	52	·097	71	$\cdot 0^{4}68$		
		40	$\cdot 0^{3}58$	53	·073	72	$\cdot 0^{4}48$		
		45	$\cdot 0^{4}31$	54	·057	73	$\cdot 0^{4}16$		

to a frequency $\binom{m}{\gamma}$ is $\binom{m-\gamma}{2}+\binom{\gamma}{2}$, which is a quadratic in γ. Were the variate value a linear function of γ the distribution for the single cell would tend to normality in accordance with the well-known property of the binomial. The case of the quadratic value corresponds to a transformation of the variate of the type $x^2=y$ and the transform of the normal form $\exp(-x^2)\,dx$ becomes the Type III form $\exp(-y)\,y^{-\frac{1}{2}}dy$. Since the N cells are independent and the sum of variates in the same Type III form is also distributed in that form, it follows that Σ is in the limit distributed as $\exp(-\Sigma)\,\Sigma^{\frac{N}{2}-1}\,d\Sigma$ except perhaps for some constants. Thus Σ or some multiple of it is distributed as χ^2.

For constant m the distribution tends to normality with increasing n.

18. The first of these results suggests that the Type III distribution will provide an approximation to the distribution (13) when m is moderately large. We proceed to find the first four moments of (13).

It is sufficient to find the first four moments of (12), those of (13) being obtainable therefrom in virtue of the relationships which connect seminvariants of independent distributions.

The rth moment of (12) about the origin is given by

$$2^m\mu_r'=\left[\left(t\frac{\partial}{\partial t}\right)^r f\right]_{t=1}, \qquad \text{......(14)}$$

since 2^m is the total frequency. Thus we have

$$2^m\mu_1'=\mathop{S}_{r=0}^{m}\binom{m}{r}\left(r^2-mr+\frac{m^2-m}{2}\right)=2^m\binom{m}{2}+S\binom{m}{r}(r^2-mr). \qquad \text{......(15)}$$

Sums such as $S\binom{m}{r}r^p$ can be obtained by operating on the binomial $(1+x)^m$ p times by $x\frac{\partial}{\partial x}$, e.g. we find

$$S\left\{\binom{m}{r}r\right\}=2^m\frac{m}{2},$$

$$S\binom{m}{r}r^2=2^m\left\{\frac{m}{2}+\frac{1}{2}\binom{m}{2}\right\}$$

and hence, substituting in (15),

$$\mu_1'=\frac{1}{2}\binom{m}{2}.$$

Thus the mean of the distribution (13) is given by

$$\mu_1'=\tfrac{1}{2}N\binom{m}{2}. \qquad \text{......(16)}$$

In a similar way we find

$$\mu_2 = \tfrac{1}{4}N\binom{m}{2}, \qquad \text{......(17)}$$

$$\mu_3 = \tfrac{3}{4}N\binom{m}{3}, \qquad \text{......(18)}$$

$$\mu_4 = N\binom{m}{2}\left\{\frac{3m^2-15m+17}{8} + \tfrac{3}{32}N(m^2-m)\right\}. \qquad \text{......(19)}$$

These are the moments of Σ. Those of u are obtained by dividing by an appropriate power of $N\binom{m}{2}$ and it may be noted in particular that the mean of u is zero.

We have directly from (17), (18) and (19)

$$\beta_1 = \frac{8}{N}\frac{(m-2)^2}{m(m-1)},$$

$$\beta_2 = \frac{4}{Nm(m-1)}\left\{3m^2-15m+17+\frac{3N}{4}m(m-1)\right\}.$$

For constant m, as $N \to \infty$,

$$\beta_1 \to 0, \quad \beta_2 \to 3$$

and for constant N, as $m \to \infty$,

$$\beta_1 \to \frac{8}{N}, \quad \beta_2 \to \frac{12}{N}+3,$$

confirming the tendency towards the Type III distribution.

19. The first four moments of the Type III distribution

$$dF = ke^{-px}\,x^{q-1}\,dx$$

are

$$\frac{q}{p}, \frac{q}{p^2}, \frac{2q}{p^3}, \frac{3q(q+2)}{p^4}.$$

Equating the second and third moments to those given by (17) and (18) we find

$$q = \frac{Nm(m-1)}{2(m-2)^2}, \qquad \text{......(20)}$$

$$p = \frac{2}{m-2}. \qquad \text{......(21)}$$

To make the first moments correspond we move the origin of the Σ dis-

tribution a distance $\frac{1}{2}N\binom{m}{2}\frac{m-3}{m-2}$ to the right. We thus reach the approximation to the Σ distribution, coinciding in the first three moments

$$dF = ke^{-\frac{2x}{m-2}}x^{\frac{Nm(m-1)}{2(m-1)^2}-1}dx,$$

where

$$x = \Sigma - \tfrac{1}{2}N\binom{m}{2}\frac{m-3}{m-2}$$

or, transforming to the more usual χ^2 form by putting $\chi^2 = 4x/(m-2)$, we find that

$$\left\{\Sigma - \tfrac{1}{2}N\binom{m}{2}\frac{m-3}{m-2}\right\}\frac{4}{m-2} \qquad \ldots\ldots(22)$$

is distributed as χ^2 with

$$\nu = \frac{Nm(m-1)}{(m-2)^2} \qquad \ldots\ldots(23)$$

degrees of freedom.

The fourth moments of Σ and the χ^2 approximation differ by terms of order N^{-1} and m^{-1} compared with their absolute values.

20. It only remains to be seen how large m and n must be for this to provide a satisfactory approximation.

Consider first the distributions for $m = 3$. When $n = 8$, $N = 28$, we have, for the approximation, 4Σ distributed with 168 degrees of freedom. From Table III we see that for $\Sigma = 54$, $P = 0{\cdot}011$ and for $\Sigma = 58$, $P = 0{\cdot}0011$. Applying a continuity correction by deducting unity from Σ we find for the χ^2 approximation with $\chi^2 = 4\times 53$, $\nu = 168$, $P = 0{\cdot}011$, and with $\chi^2 = 4\times 57$, $P = 0{\cdot}00114$. The correspondence is very close, in spite of the low value of m.

For $m = 4$, $n = 5$, $N = 10$, the approximation gives $2\Sigma - 30$ distributed with 30 degrees of freedom. For $\Sigma = 40$ and 41, this gives, with continuity corrections of 0·5, half the variate-interval, $\chi^2 = 49$ and 51, $\nu = 30$. From the diagram given in Yule & Kendall's "Introduction to the Theory of Statistics" (1937) it is seen that these values lie one on either side of the 1% value; and this is in accordance with the exact values of P, which are seen from Table IV to be 0·016 and 0·0088. Similarly we find that the values of Σ, 37 and 38, lie on either side of the 5% level, which is again in accordance with the exact values, $P = 0{\cdot}060$ and 0·038.

For $m = 6$, $n = 4$, $N = 6$, the approximation gives $\Sigma - 33{\cdot}75$ distributed with 11·25 degrees of freedom. For $\Sigma = 59$ and 60 the corresponding χ^2 values are seen to lie on either side of the 1% point, which accords with the exact value of Table VI.

We conclude that the χ^2 approximation provides an adequate test of significance for the values of m and n outside the range for which Tables III–VI give exact values.

21. As a matter of theoretical interest we may record the results for the distribution of u when the data are ranked. It appears that in this case

$$\frac{12}{m-2}\frac{2n+5}{2n^2+6n+7}\left[\Sigma-\frac{1}{2}\binom{m}{2}\binom{n}{2}\left\{1-\frac{1}{3(m-2)}\frac{(2n+5)^2}{2n^2+6n+7}\right\}\right]$$

is distributed approximately as χ^2 with

$$\binom{m}{2}\frac{2}{(m-2)^2}\binom{n}{2}\frac{(2n+5)^3}{(2n^2+6n+7)}\text{ degrees of freedom.}$$

This result is not of much practical value. The case of m rankings can be more simply treated by other methods.

INTERPRETATION OF RESULTS OF PAIRED COMPARISONS

22. In the light of the foregoing theory we may discuss the interpretation of the results of a paired comparison experiment.

If for each observer the coefficient of consistence is unity the comparisons reduce to rankings and may be discussed by known methods. But if some or all the coefficients are not unity we have to consider the following possibilities:

(*a*) Some of the observers may be bad judges and the inconsistences reflect their shortcomings in making comparisons.

(*b*) Some of the objects may differ by amounts which fall below the threshold of distinguishability for some observers.

(*c*) The property under judgment may not be a linear variate at all and we may be getting the sort of confusion which would result if observers were asked to compare English towns according to the bivariate concept "gēographical position".

(*d*) Several of the effects may be operating simultaneously.

23. If we have only one observer and have no prior knowledge of his capabilities it is not in general possible to apportion his inconsistences among these causes. Exceptions may occur when the inconsistences are of a marked and peculiar kind; for instance if they involve only four objects out of 15, we may suspect that the four are practically indistinguishable rather than that the observer is unable to make distinctions at all and avoided inconsistences among the others by sheer chance. But even here conclusions drawn *a posteriori* after inspection of the data are dubious. Table II gives a test of the hypothesis that an observer is incapable of making judgments. For example, with $n = 7$, the chances are 983 in 1000 that if the preferences are made at random there will be more than two inconsistent triads, so that if we find two or less, it is improbable that the observer is completely incapable of judgment. We might then be led to suppose that his small deviation from internal consistence is due to fluctuation of attention, very close resemblance to the objects giving rise to the inconsistences, or both.

24. With m observers the investigation can be taken a good deal further. If all the observers show inconsistences we suspect that the objects are at fault or that the observers are being asked to perform an impossible task. On the other hand, if most of the observers show a small or zero inconsistence we suspect that the others are just bad judges and may reject their data accordingly.

As between indistinguishability of objects and non-linearity of variate, a choice of explanations would depend largely on the extent to which inconsistences were concerned with the same set of objects. If there is a high value of u, indicating concordance of judgment, we expect to find most of the inconsistences confined to certain objects, and common to observers. In this case we suspect that the objects are close together in the degree to which they exhibit the quality under consideration. But if the observers scatter their inconsistences over the whole field u will be moderate or low and we suspect that the observers are being asked to do something beyond their capacity; and this brings us to question the validity of regarding the quality as a linear variable.

25. When a quality such as "bravery" or "intelligence" is insusceptible to measurement there is frequently doubt of this kind. But this has not deterred investigators from assuming that such statistical variables exist, or from requiring observers to rank objects according to them, or in some cases from replacing such rankings by quantiles of the normal curve. We are never tired of criticizing this Principle of the Hypostasis of Plausible Terminology. Hitherto it has flourished largely because of the difficulty of adducing evidence against it; and we hope that the inconsistence of paired comparisons will provide a criterion, however rough, of the legitimacy of the methods to which it leads.

But we would emphasize that our approach to the method of paired comparisons has a somewhat different object from that elaborated by Thurstone (1927 and many subsequent papers). As we understand it, his method is appropriate where one is entitled to assume *a priori* or by reason of precautions taken in the selection of material that a linear variable is involved and that there exist perceptible differences between the items presented for comparison. Our object is to make it possible to dispense with such assumptions and precautions.

26. A few words may be added about the case in which an objective order is known to exist (as, for instance, in judging individuals according to age or weight). In such circumstances the appearance of inconsistences will indicate unreliability of the part of the observer or subliminal differences between objects. A measure of the observer's reliability may be obtained by calculating u between known and observed comparisons. If ζ is high enough to enable us to accept his judgments as internally consistent on the whole, u may still be low enough to reject his judgments as accurate.

27. We conclude with an example of the application of the foregoing theory to some experimental material.

Classes of children (ages 11 to 13 inclusive) were asked to state their preferences with respect to certain school subjects. Each child was given a sheet on which were written the possible pairs of subjects and asked to underline the one preferred in each case. Two classes gave the following results:

(*a*) 21 boys, 13 school subjects. The preferences are shown in Table VII, which is in the form described in section 13; e.g. there were 18 boys who preferred Art to Religion.

TABLE VII

Preferences of 21 boys in 13 subjects

	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
1. Woodwork	—	14	20	15	15	16	16	18	18	18	20	21	20	211
2. Gymnastics	7	—	14	12	13	18	14	16	16	20	16	18	19	183
3. Art	1	7	—	10	14	10	16	18	16	16	17	16	19	160
4. Science	6	9	11	—	11	12	15	14	13	13	17	17	16	154
5. History	6	8	7	10	—	14	11	12	14	15	13	14	16	140
6. Geography	5	3	11	9	7	—	14	14	13	13	16	15	17	137
7. Arithmetic	5	7	5	6	10	7	—	9	11	13	15	13	15	116
8. Religion	3	5	3	7	9	7	12	—	12	14	14	16	14	116
9. English Literature	3	5	5	8	7	8	10	9	—	10	13	13	15	106
10. Commercial subjects	3	1	5	8	6	8	8	7	11	—	10	10	14	91
11. Algebra	1	5	4	4	8	5	6	7	8	11	—	10	13	82
12. English Grammar	0	3	5	4	7	6	8	5	8	11	11	—	13	81
13. Geometry	1	2	2	5	5	4	6	7	6	7	8	8	—	61
													Total	1638

The calculation of Σ for this table, in which the objects are arranged in order of total number of preferences, may be shortened by noting that Σ as given by equation (7) may be transformed into the form

$$\Sigma = S(\gamma^2) - mS(\gamma) + \binom{m}{2}\binom{n}{2},$$

where the summation now takes place over the half of the table below the diagonal. Since the numbers in this half are smaller than those in the other half there is a considerable saving in arithmetic.

We find $$\Sigma = 9718$$

and hence $$u = \frac{2 \times 9718}{\binom{21}{2}\binom{13}{2}} - 1 = 0{\cdot}186.$$

There is thus a certain amount of agreement among the children, indicated by the positive value of u. Is this significant?

We note first of all that this distribution of preferences could not have arisen by chance to any acceptable degree of probability. In fact, $\chi^2 = 412{\cdot}4$ (equation (22)) and $\nu = 90{\cdot}7$. The large value of ν justifies the use of the normal

approximation to the χ^2 distribution and we find $\sqrt{(2\chi^2)} - \sqrt{(2\nu - 1)} = 15{\cdot}3$ a very improbable result on the hypothesis of a random allocation of preferences.

The distribution of circular triads was as follows:

No. of triads	Frequency	No. of triads	Frequency
0	1	12	1
1	1	17	3
4	5	21	1
6	2	25	1
7	2	29	1
8	1	30	1
10	1	Total	21

The total number of circular triads was 242 with a mean of 11·5. Only one boy was entirely consistent. On the other hand, for $n = 13$ the maximum number of circular triads is 91, with a mathematical expectation of 71·5. It is thus clear that, except perhaps for one boy, we cannot suppose that any boy allotted preferences at random. We are again led to conclude that the boys are genuinely capable of making distinctions, and that consistently on the whole. Half the boys have coefficients of consistence ζ greater than 0·92.

We conclude that the boys can make preferences and that in their view the subjects are sufficiently different to enable a reasonably consistent set of preferences to be made. So far as these data are concerned we would see no objection to the assumption that a *scale* of preferences can be set up. With this in mind we can say that the value of u indicates a certain amount of agreement, though not a strong one, between the boys as to which subjects they prefer.

(*b*) 25 girls, 11 school subjects. Table VIII shows the data.

TABLE VIII

Preferences of 25 girls in 11 subjects

	1	2	3	4	5	6	7	8	9	10	11	Total
1. Gymnastics	—	10	19	17	20	17	21	21	21	18	22	186
2. Science	15	—	12	15	17	15	21	19	18	16	17	165
3. Art	6	13	—	16	16	18	10	17	16	10	16	147
4. Domestic Science	8	10	9	—	16	11	13	15	14	11	14	121
5. History	5	8	9	9	—	14	18	12	13	15	18	121
6. Arithmetic	8	10	7	14	11	—	12	13	12	16	18	121
7. Geography	4	4	15	12	7	13	—	14	15	14	14	112
8. English Literature	4	6	8	10	13	12	11	—	14	13	14	105
9. Religion	4	7	9	11	12	13	10	11	—	11	17	105
10. Algebra	7	9	6	14	10	9	11	12	14	—	12	104
11. English Grammar	3	8	9	11	7	7	11	11	8	13	—	88
											Total	1375

We find $\Sigma = 8928$, $u = 0{\cdot}082$.

For the χ^2 significance test, $\chi^2 = 180{\cdot}3$, $\nu = 62{\cdot}4$ and $\sqrt{(2\chi^2)} - \sqrt{(2\nu - 1)} = 7{\cdot}9$, as before a very significant result.

The distribution of circular triads was

No. of triads	Frequency	No. of triads	Frequency
1	2	17	1
2	2	19	1
3	1	22	1
4	1	23	2
6	1	27	1
8	1	32	1
9	1	35	1
11	2	37	1
12	2	38	1
13	1		—
14	1	Total	25

The total number of circular triads is 382 with a mean 15·28. For $n = 11$ the maximum number of circular triads is 55 with an expectation of 41·25. Several of the girls come very close to this, the worst having a coefficient of consistence equal to 0·31.

We are, however, again led to conclude that the preferences were not allotted at random and that most of the girls are capable of exercising a judgment which is on the whole consistent. There is only a very slight agreement in preferences.

Thus the girls are less consistent and less alike in preferences than the boys.

REFERENCES

KENDALL, M. G. (1938). "A New Measure of Rank Correlation." *Biometrika*, **30**, 81.

KENDALL, M. G. & BABINGTON SMITH B. (1939). "On the Problem of *m* Rankings." *Ann. Math. Statist.* **10**, 273.

THURSTONE, L. L. (1927). "A Law of Comparative Judgment." *Psychol. Rev.* **34**, 275.

THE MEAN AND VARIANCE OF χ^2, WHEN USED AS A TEST OF HOMOGENEITY, WHEN EXPECTATIONS ARE SMALL

By J. B. S. HALDANE, F.R.S.

PEARSON's measure of divergence, χ^2, can be used not only as a test of goodness of fit, but as a test of homogeneity when the expectations are unknown before the samples are observed. Consider the $(m \times n)$-fold table:

$$
\begin{array}{|ccccc|c|}
a_{11} & a_{12} & a_{13} & \dots & a_{1m} & s_1 \\
a_{21} & a_{22} & a_{23} & \dots & a_{2m} & s_2 \\
a_{31} & a_{32} & a_{33} & \dots & a_{3m} & s_3 \\
\dots & \dots & \dots & \dots & \dots & \dots \\
\dots & \dots & \dots & \dots & \dots & \dots \\
a_{n1} & a_{n2} & a_{n3} & \dots & a_{nm} & s_n \\
\hline
t_1 & t_2 & t_3 & \dots & t_m & N
\end{array}
$$

Here each a_{ij} represents a number of individuals observed, and $s_i = \sum_j a_{ij}$, $t_j = \sum_i a_{ij}$, $N = \sum_i s_i = \sum_j t_j$. The table may represent n samples of $s_1, s_2, \dots, s_n$ individuals, each sample falling into m classes, $t_1, t_2, \dots, t_m$ being the grand totals in each class. Or it may represent m samples of $t_1, t_2, \dots, t_m$ individuals, each falling into n classes, the class totals being $s_1, s_2, \dots, s_n$. In each case we ask what is the probability of so bad a fit if every sample is taken from the same large population. The expected value of a_{ij} is clearly $\frac{s_i t_j}{N}$, and

$$\chi^2 = \sum_{ij} \frac{\left(a_{ij} - \frac{s_i t_j}{N}\right)^2}{\frac{s_i t_j}{N}}$$

$$= N \sum_{ij} \frac{(a_{ij}^2)}{(s_i t_j)} - N.$$

Fisher (1922) showed that when every s_i and t_j was sufficiently large, χ^2 has the usual distribution, with $(m-1)(n-1)$ degrees of freedom. Thus its mean is $(m-1)(n-1)$, its variance $2(m-1)(n-1)$. Haldane (1937, 1938, 1939) investigated the exact values of the moments of χ^2 when expectations are small, in $(m \times n)$-fold tables with mn or $m(n-1)$ degrees of freedom, but did not attempt the present problem. This had previously been done by Cochran (1936), in the

restricted case when $m = 2$, and every s_i is equal. Cochran used a method based on the use of characteristic functions. My own method is entirely elementary, and completely exact, though rather tedious. My results are very close to Cochran's, but the accurate values are worth putting on record.

THE MEAN VALUE OF χ^2

Given the marginal totals, the probability of any particular distribution is

$$P = \frac{\prod_i (s_i!) \prod_j (t_j!)}{N! \prod_{ij} (a_{ij}!)}.$$

And this is so whatever may be the true expectations. Thus, if in the sample s_i the expectation of the observation a_{ij} is $s_i p_j$, so that that of t_j is Np_j, then the probability of the given distribution is

$$\frac{\prod_j (p_j^{t_j}) \prod_i (s_i!)}{\prod_{ij} (a_{ij}!)}.$$

But the probability of the given marginal totals t_j is

$$\frac{\prod_j (p_j^{t_j}) N!}{\prod_j (t_j!)}.$$

Hence P is the probability of obtaining any particular distribution with the given sample sizes s_i and class totals t_j.

If $E(x)$ denotes the sampling expectation of x, then $E(a_{ij}) = \Sigma\, Pa_{ij}$, summation being taken over all samples possible with the given marginal totals. Hence

$$E(a_{ij}) = \frac{s_i t_j}{N} \sum \frac{(s_i-1)!\,(t_j-1)! \prod_k (s_k!) \prod_l (t_l!)}{(N-1)!\,(a_{ij}-1)! \prod_{kl} (a_{kl}!)},$$

where k assumes all positive integral values between 0 and n except i, and l all positive integral values between 0 and m except j. That is to say the quantity summed is the probability of a set of observations similar to those of the table except that s_i and t_j have been diminished by unity. Hence the sum equals unity.

Similarly
$$E[a_{ij}(a_{ij}-1)] = \frac{s_i(s_i-1)\,t_j(t_j-1)}{N(N-1)},$$

$$E(a_{ij}a_{kj}) = \frac{s_i s_k t_j(t_j-1)}{N(N-1)},$$

$$E(a_{ij}a_{kl}) = \frac{s_i s_k t_j t_l}{N(N-1)},$$

and so on. Hence

$$E(\chi^2) = \overline{\chi^2} = NE\left[\sum_{ij}\frac{(a_{ij}^2)}{(s_i t_j)}\right] - N$$

$$= N\sum_{ij}\left[\frac{1}{s_i t_j}E(a_{ij}^2)\right] - N$$

$$= \sum_{ij}\left[\frac{(s_i-1)(t_j-1)}{N-1}+1\right] - N$$

$$= \frac{1}{N-1}\sum_{ij}(s_i t_j - s_i - t_j + N) - N$$

$$= \frac{1}{N-1}(N^2 - nN - mN + mnN) - N$$

$$= \frac{(m-1)(n-1)N}{N-1}, \qquad \ldots\ldots(1)$$

a result previously obtained by Bartlett (1937).

This clearly becomes $(m-1)(n-1)$ when N tends to infinity. In the case where $m=2$, and every $s_i = s$, we have

$$\overline{\chi^2} = \frac{(n-1)ns}{ns-1}. \qquad \ldots\ldots(2)$$

Cochran gave
$$\overline{\chi^2} = n-1+\frac{ns-s+1}{ns^2}.$$

This exceeds (2) by $\frac{s-1}{ns^2(ns-1)}$, and is therefore correct to $0(n^{-1})$. It will be noted that the mean depends only on the number of degrees of freedom and the grand total N. The higher moments involve the marginal totals s_i and t_j, and are therefore more complicated. It is clear that the expectation is diminished because, in sampling from a finite group of N individuals, $E(a_{ij})$ has the same value as in sampling from an infinite population with the same frequency, but $E(a_{ij}^2)$ is less.

The variance in a $(2\times n)$-fold table with equal samples

Owing to the complexity of the general expression for the variance, we shall first calculate it for the $(2\times n)$-fold table:

a_1	a_2	a_3	...	a_n	A
b_1	b_2	b_3	...	b_n	B
s	s	s	...	s	N

Here n samples of s are taken, and each falls into two classes, say a_i healthy and b_i diseased. The totals are A and B, and $N = ns = A+B$. It will be convenient to put $A = pN$, $B = qN$.

$$\chi^2 = \frac{1}{spq}\Sigma(sp-a_i)^2$$

$$= \frac{1}{spq}\Sigma a_i^2 - \frac{nps}{q},$$

$$\left(\chi^2 + \frac{nps}{q}\right)^2 = \frac{1}{s^2p^2q^2}(\Sigma a_i^2)^2.$$

Now
$$(\Sigma a_i^2)^2 = \sum_i a_i^4 + 2\sum_{ij} a_i^2 a_j^2,$$

summation being taken over all unequal pairs of values of i and j.
Hence

$$E(\Sigma a_i^2)^2 = nE(a_i^4) + n(n-1)\,E(a_i^2 a_j^2)$$

$$= nE[a_i(a_i-1)(a_i-2)(a_i-3) + 6a_i(a_i-1)(a_i-2) + 7a_i(a_i-1) + a_i]$$
$$+ n(n-1)\,E[a_i(a_i-1)\,a_j(a_j-1) + 2a_i(a_i-1)\,a_j + a_i a_j]$$

$$= \frac{A(A-1)(A-2)(A-3)}{N(N-1)(N-2)(N-3)}[ns(s-1)(s-2)(s-3) + n(n-1)\,s^2(s-1)^2]$$

$$+ \frac{A(A-1)(A-2)}{N(N-1)(N-2)}[6ns(s-1)(s-2) + 2n(n-1)\,s^2(s-1)]$$

$$+ \frac{A(A-1)}{N(N-1)}[7ns(s-1) + n(n-1)\,s^2] + \frac{A}{N}ns.$$

Since $N = ns$, $A = pns$, we have

$$\frac{1}{pns}E(\Sigma a_i^2)^2 = \frac{(A-1)(A-2)(A-3)}{(N-1)(N-2)(N-3)}(s-1)[ns^2 - (n+4)s + 6]$$

$$+ \frac{2(A-1)(A-2)}{(N-1)(N-2)}(s-1)[(n+2)s - 6]$$

$$+ \frac{A-1}{N-1}[(n+6)s - 7] + 1,$$

$$E(\Sigma a_i^2)^2 = \frac{pn^2s^3}{(ns-1)(ns-2)(ns-3)}[p^3n^3s^3(s-1)\{ns^2 - (n+4)s + 6\}$$

$$+ 2p^2n(n-1)\,s(s-1)(ns-6)$$

$$+ p(n-1)\{n(n+1)\,s^2 - 2(3n-4)\,s - 2\} - 2(n-1)(s-1)].$$

So the variance of χ^2 is given by

$$V(\chi^2) = \frac{E(\Sigma a_i^2)^2}{p^2q^2s^2} - \left[\frac{(n-1)\,ns}{(ns-1)} + \frac{pns}{q}\right]^2$$

$$= \frac{E(\Sigma a_i^2)^2}{p^2q^2s^2} - \frac{n^2s^2[pn(s-1)+n-1]^2}{q^2(ns-1)^2}$$

$$= \frac{n^2s}{pq^2(ns-1)^2\,(ns-2)\,(ns-3)}\,[(ns-1)\{p^3n^2s(s-1)\,(ns^2-\overline{n-4}s+6)$$

$$+2p^2n(n-1)\,s(s-1)\,(ns-6)$$

$$+p(n-1)\,(n\overline{n+1}s^2-6\overline{n-8}s-2)-2(n-1)\,(s-1)\}$$

$$-ps(ns-2)\,(ns-3)\,\{pn(s-1)+n-1\}^2]$$

$$= \frac{2n^2(n-1)\,s(s-1)}{pq^2(ns-1)^2\,(ns-2)\,(ns-3)}\,[p^3n^2s^2-2p^2n^2s^2+p(n^2s^2+ns-1)-(ns-1)]$$

$$= \frac{2n^2(n-1)\,s(s-1)\,(pqn^2s^2-ns+1)}{pq(ns-1)^2\,(ns-2)\,(ns-3)}$$

$$= \frac{2n^4(n-1)\,s^3(s-1)}{(ns-1)^2\,(ns-2)\,(ns-3)}\left(1-\frac{ns-1}{AB}\right)$$

$$= \frac{2(n-1)\,N^3(N-n)}{(N-1)^2\,(N-2)\,(N-3)}\left(1-\frac{N-1}{AB}\right). \qquad \ldots\ldots(3)$$

This becomes $2(n-1)$ when A and B tend to infinity.

When A and B are both of order N, that is to say neither p nor q is small, it becomes

$$2(n-1)\left[1-\left(n+\frac{1}{pq}-7\right)N^{-1}\right]+0(N^{-2}).$$

Cochran's expression is $2(n-1)\,(1-nN^{-1})+0(N^{-2})$, which is accurate when $p = \frac{1}{2} \pm \frac{\sqrt{21}}{14}$, i.e. 0·8273 or 0·1727. It will be seen that the variance is diminished when $n > 7-\frac{1}{pq}$, which is certainly the case if $n > 3$. If A remains finite when N and B tend to infinity we find $\overline{\chi^2} = n-1$,

$$V(\chi^2) = 2(n-1)\left(1-\frac{1}{A}\right),$$

a formula already given by Haldane (1937).

Formula (3) may be compared with the values given by Haldane (1937) for

a $(n \times 2)$-fold table with n degrees of freedom. In the terminology of this paper they are:

$$\overline{\chi^2} = n,$$

$$V(\chi^2) = 2n + \frac{n}{s}\left(\frac{1}{pq} - 6\right).$$

Similarly the limiting case given above may be compared with the values for a n-fold table with n degrees of freedom, namely,

$$\overline{\chi^2} = n,$$

$$V(\chi^2) = 2n + \frac{n}{a}.$$

THE VARIANCE FOR A $(m \times n)$-FOLD TABLE

This of course includes the variance found above as a special case. However, as the summations involved are somewhat hard to follow, the simpler algebra for the special case has been given separately.

$$\chi^2 = N \sum_{ij}\left(\frac{a_{ij}^2}{s_i t_j}\right) - N.$$

Hence $\quad N^{-2}(\chi^2 + N)^2 = \left[\sum_{ij}\left(\frac{a_{ij}^2}{s_i t_j}\right)\right]^2$

$$= \sum_{ij}\left(\frac{a_{ij}^4}{s_i^2 t_j^2}\right) + \sum_{ikj}\left(\frac{a_{ij}^2 a_{kj}^2}{s_i s_k t_j^2}\right) + \sum_{ijl}\left(\frac{a_{ij}^2 a_{il}^2}{s_i^2 t_j t_l}\right) + \sum_{ijkl}\left(\frac{a_{ij}^2 a_{kl}^2}{s_i s_j t_j t_l}\right).$$

Here summation over the values of i, k on the one hand, and j, l on the other, are independent and commutative. $\sum_{ik}(x) = \sum_i\left[\sum_k(x)\right]$, where k assumes all values from o to n inclusive, except i. $\sum_{jl}(x)$ has a similar meaning.

Thus $\sum_i(1) = n$, $\sum_i(s_i) = N$, $\sum_i\left[1 + \sum_k(1)\right] = n^2$, $\sum_i\left[s_i + \sum_k s_k\right] = nN$, and so on.

Let $$S = \sum_i s_i^{-1},\ T = \sum_j t_j^{-1}.$$

Then

$$a_{ij}^4 = a_{ij}(a_{ij}-1)(a_{ij}-2)(a_{ij}-3) + 6a_{ij}(a_{ij}-1)(a_{ij}-2) + 7a_{ij}(a_{ij}-1) + a_{ij},$$

$$a_{ij}^2 a_{kj}^2 = a_{ij}(a_{ij}-1)a_{kj}(a_{kj}-1) + a_{ij}(a_{ij}-2)a_{kj} + a_{ij}a_{kj}(a_{kj}-1) + a_{ij}a_{kj},$$

and so on. The expectations of the two middle terms in the last expression are of course equal. Remembering that

$$E[a_{ij}(a_{ij}-1)(a_{ij}-2)(a_{ij}-3)] = \frac{s_i(s_i-1)(s_i-2)(s_i-3)t_j(t_j-1)(t_j-2)(t_j-3)}{N(N-1)(N-2)(N-3)},$$

and $\quad E[a_{ij}(a_{ij}-1)a_{kj}(a_{kj}-1)] = \dfrac{s_i(s_i-1)s_k(s_k-1)t_j(t_j-1)(t_j-2)(t_j-3)}{N(N-1)(N-2)(N-3)},$

and so on, we have

$$N^{-2}E[(\chi^2+N)^2]$$

$$= \frac{1}{N(N-1)(N-2)(N-3)}\left[\sum_{ij}\frac{(s_i-1)(s_i-2)(s_i-3)(t_j-1)(t_j-2)(t_j-3)}{s_it_j}\right.$$

$$+\sum_{ikj}\frac{(s_i-1)(s_k-1)(t_j-1)(t_j-2)(t_j-3)}{t_j}$$

$$\left.+\sum_{ijl}\frac{(s_i-1)(s_i-2)(s_i-3)(t_j-1)(t_l-1)}{s_i}+\sum_{ikjl}(s_i-1)(s_k-1)(t_j-1)(t_l-1)\right]$$

$$+\frac{1}{N(N-1)(N-2)}\left[6\sum_{ij}\frac{(s_i-1)(s_i-2)(t_j-1)(t_j-2)}{s_it_j}\right.$$

$$+2\sum_{ikj}\frac{(s_i-1)(t_j-1)(t_j-2)}{t_j}+2\sum_{ijl}\frac{(s_i-1)(s_i-2)(t_j-1)}{s_i}$$

$$\left.+2\sum_{ikjl}(s_i-1)(t_j-1)\right]$$

$$=\frac{1}{N(N-1)}\left[7\sum_{ij}\frac{(s_i-1)(t_j-1)}{s_it_j}+\sum_{ikj}\frac{t_j-1}{t_j}+\sum_{ijl}\frac{s_i-1}{s_i}+\sum_{ikjl}(1)\right]+\frac{1}{N}\sum_{ij}\left(\frac{1}{s_it_j}\right).$$

Hence

$$N^{-1}E[(\chi^2+N)^2]$$

$$=\frac{1}{(N-1)(N-2)(N-3)}\left[\sum_i(s_i^2-6s_i+11-6s_i^{-1})+\sum_{ik}(s_i-1)(s_k-1)\right]$$

$$\times\left[\sum_j(t_j^2-6t_j+11-6t_j^{-1})+\sum_{jl}(t_j-1)(t_l-1)\right]$$

$$+\frac{2}{(N-1)(N-2)}\left[2\sum_i(s_i-3+2s_i^{-1})\sum_j(t_j-3+2t_j^{-1})\right.$$

$$\left.+\left\{\sum_i(s_i-3+2s_i^{-1})+\sum_{ik}(s_i-1)\right\}\left\{\sum_j(t_j-3+2t_j^{-1})+\sum_{jl}(t_j-1)\right\}\right]$$

$$+\frac{1}{N-1}\left[6\sum_i(1-s_i^{-1})\sum_j(1-t_j^{-1})+\left\{\sum_i(1-s_i^{-1})+\sum_{ik}(1)\right\}\right.$$

$$\left.\times\left\{\sum_j(1-t_j^{-1})+\sum_{jl}(1)\right\}\right]+\frac{1}{N}\sum_i(s_i^{-1})\sum_j(t_j^{-1})$$

$$=\frac{1}{(N-1)(N-2)(N-3)}[N^2-2(n+2)N+n^2+10n-6S]$$

$$\times[N^2-2(m+2)N+m^2+10m-6T]+\frac{2}{(N-1)(N-2)}$$

$$\times[2(N-3n+2S)(N-3m+2T)+(nN-n^2-2n+2S)(mN-m^2-2m+2T)]$$

$$+\frac{1}{N-1}[6(n-S)(m-T)+(n^2-S)(m^2-T)]+ST.$$

Hence

$$N^{-1}(N-1)(N-2)(N-3)\,E[(\chi^2+N)^2]$$

$$\begin{aligned}
&= [N^2-2(n+2)N+n^2+10n-6S]\,[N^2-2(m+2)N+m^2+10m-6T] \\
&\quad +2(N-3)\,[(nm+2)N^2 \\
&\quad -\{nm(n+m)+4nm+6(n+m)-2(n+2)T-2(m+2)S\}N \\
&\quad +n^2m^2+2nm(n+m)+22nm-2(n^2+8n)T-2(m^2+8m)S+12ST] \\
&\quad +(N-2)(N-3)\,[n^2m^2+6nm-(n^2+6n)T-(m^2+6m)S+7ST] \\
&\quad +(N-1)(N-2)(N-3)\,ST \\
&= N^4+2(\alpha-\beta-2)N^3+[(\alpha-\beta)^2-6\alpha+6\beta+4]N^2 \\
&\quad +(-3\alpha^2+8\alpha\beta-4\beta^2+6\alpha-4\beta)N+\alpha^2-2\alpha\beta+4\alpha-[(n^2+2n-2)T \\
&\quad +(m^2+2m-2)S]N^2+[(n^2-2n)T+(m^2-2m)S]N+N^2(N+1)ST,
\end{aligned}$$

where $\alpha = nm$, $\beta = n+m$.

But
$$E(\chi^2+N) = \frac{(n-1)(m-1)N}{N-1}+N = \frac{N(N+\alpha-\beta)}{N-1}.$$

Hence

$$N^{-1}(N-1)^2(N-2)(N-3)\,V(\chi^2)$$

$$\begin{aligned}
&= (N-1)\,[N^4+2(\alpha-\beta-2)N^3+\text{etc.}]-N(N-2)(N-3)(N+\alpha-\beta)^2 \\
&= 2(\alpha-\beta+1)N^3+(\alpha^2+2\beta-4)N^2-2(\alpha^2-\alpha\beta+\beta^2+\alpha-2\beta)N \\
&\quad -\alpha(\alpha-2\beta+4)-[(n^2+2n-2)T+(m^2+2m-2)S]N^2(N-1) \\
&\quad +[(n^2-2n)T+(m^2-2m)S]N(N-1)+STN^2(N^2-1).
\end{aligned}$$

Thus

$$\begin{aligned}
V(\chi^2) &= \frac{N}{(N-1)(N-2)(N-3)} \\
&\times\Bigg[\frac{\begin{array}{l}2(n-1)(m-1)N^3+(n^2m^2+2n+2m-4)N^2 \\ \qquad -2\{nm(n-1)(m-1)+(n+m)(n+m-2)\}N-nm(n-2)(m-2)\end{array}}{N-1} \\
&\quad -\{(n^2+2n-2)T+(m^2+2m-2)S\}N^2 \\
&\qquad +\{n(n-2)T+m(m-2)S\}N+STN^2(N+1),\Bigg] \qquad \ldots\ldots(4)
\end{aligned}$$

where S and T have been defined on p. 351 above.

When there are n samples of $s_i, s_2, s_3, \ldots, s_n$ members, falling into two classes whose totals are A and B, we have $m = 2$, $T = \frac{N}{AB}$, $S = \Sigma s_i^{-1}$, so

$$V(\chi^2) = \frac{N^2}{(N-1)^2(N-2)(N-3)}\Big[2(n-1)N^2 + 2n(2n+1)N - 6n^2$$

$$-6N(N-1)\Sigma s_i^{-1} + \frac{N(N-1)}{AB}\{N(N+1)\Sigma s_i^{-1} - (n^2+2n-2)N + n(n-2)\}\Big]. \quad \ldots\ldots(5)$$

When every $s_i = s$, this reduces to formula (3). In the case of the fourfold table $\begin{array}{|c|c|}\hline a & b \\ \hline c & d \\ \hline\end{array}$, it becomes:

$$V(\chi^2) = \frac{N^2}{(N-1)(N-2)(N-3)}\Big[\frac{2(N^2+10N-12)}{N-1}$$

$$-6N^2\Big\{\frac{1}{(a+b)(c+d)} + \frac{1}{(a+c)(b+d)}\Big\} + \frac{N^3(N+1)}{(a+b)(c+d)(a+c)(b+d)}\Big]. \quad \ldots(6)$$

Discussion

The expressions for the higher moments would clearly be a great deal more complicated. The above calculations have, I think, a twofold interest. They show that the loss of one degree of freedom arises from the fact that we are sampling from a finite, and not an infinite, aggregate. And they point the way towards an exact treatment of the problem of curve fitting, for which Pearson originally designed the measure χ^2. In the case of a $(n \times 2)$-fold table with $(n-1)$ degrees of freedom we are, in effect, asking whether the observations give a satisfactory fit to a horizontal straight line $y = k$, where y is the observed frequency of a type within a sample. If we were trying to fit a line $y = k + lx$, we should have $n-2$ degrees of freedom, x having a different value for each sample. If we were trying to fit a normal curve we should have $n-3$ degrees, and so on. Jeffreys (1938) has pointed out the great difficulties of curve-fitting in such cases when expectations are small. It is clear that the expected value of χ^2 in such a case is not exactly $n-3$. It may turn out to be slightly greater. Thus an investigation of the actual law of error in a particular type of observation will demand an extension of the present investigation to cases where several more degrees of freedom are lost.

REFERENCES

BARTLETT, M. S. (1937). "Properties of sufficiency and statistical tests." *Proc. Roy. Soc.* A, **160**, 268–82.

COCHRAN, W. G. (1936). "The χ^2 distribution for the binomial and Poisson series, with small expectations." *Ann. Eugen., Lond.*, **7**, 207–17.

FISHER, R. A. (1922). "On the interpretation of χ^2 from contingency tables, and on the calculation of P." *J.R. Statist. Soc.* **85**, 87–94.

HALDANE, J. B. S. (1937). "The exact value of the moments of this distribution of χ^2, used as a test of goodness of fit, when expectations are small." *Biometrika*, **29**, 133–43.

— (1938). "The first six moments of χ^2 for an n-fold table with n degrees of freedom when some expectations are small." *Biometrika*, **29**, 389–91.

— (1939). "The cumulants and moments of the binomial distribution and the cumulants of χ^2 for a $(n \times 2)$-fold table." *Biometrika* (in the Press).

JEFFREYS, H. (1938). "The law of error and the combination of observations." *Philos. Trans.* A, **237**, 231–71.

Editorial Note. The expression (3) of p. 350 above for the variance of χ^2 in a $(2 \times n)$-fold table corresponds to that given by B. L. Welch in 1938 ("On tests for homogeneity," *Biometrika*, **30**, 158, equation (28)). E. S. P.

A NOTE ON THE STATISTICAL ANALYSIS OF SENTENCE-LENGTH AS A CRITERION OF LITERARY STYLE

By C. B. WILLIAMS, Sc.D.

Department of Entomology, Rothamsted Experimental Station

Some years ago I made a number of calculations of the frequency distribution of words of different length in different books to see to what extent authors kept to a definite distribution and so perhaps might be identified by such a method. The results obtained, however, were not striking and the work was put at one side.

Mr Udny Yule (1939), however, has attacked the problem of authorship from the angle of the variation in sentence length, and this appears to be a much more fertile method of approach.

Mr Yule shows that the frequency distribution of sentence length (i.e. number of words between successive full stops) is of the skew type and by comparing in two different manuscripts, the mean, the median, quartiles and deciles he is able to produce convincing mathematical evidence on the identity or otherwise of their authorship.

Mr Yule does not comment on the skew distribution further than to state (p. 371) "they are not of the Poisson type, but of the type in which the square of the standard deviation largely exceeds the mean".

When I converted some of Yule's tables into diagrams I was struck by their general resemblance to certain skew distributions with which I have recently been dealing in some Entomological problems, and which distributions, I found, became normal and symmetrical if the logarithm of the number was taken as a basis for subdivision into groups instead of the number itself (see Williams, 1927).

I was unable to test this transformation on Yule's figures as he unfortunately does not give the original data, but only the word length of sentences in groups of five; so it was necessary to obtain some new data.

These I obtained by counting the number of words in each of 600 sentences from the following three books:

(1) G. K. Chesterton, *A Short History of England*, 1917.

(2) H. G. Wells, *The Work, Wealth and Happiness of Mankind.*

(3) G. Bernard Shaw, *An Intelligent Woman's Guide to Socialism.*

All three works deal with the exposition of somewhat similar sociological subjects and none of them are in the "conversational" style.

The selection of the sentences was randomized as follows. Each of the books is divided up into chapters, sections or both. In Chesterton's book the first 30

sentences were counted in each of the first 20 chapters. In Wells's book the first 10 sentences were counted in each chapter subdivision up to chapter VII, division 11. In Shaw's book the first 15 sentences in each of sections 1–40 were taken. In each case the greater part of the book was covered.

The original data thus obtained are shown diagrammatically in Fig. 1. Each of the distributions is of the typical skew type obtained by Yule: Shaw is the most extreme and varies from 3 words to 143; Wells is less skew and ranges from 3 to 91; while the Chesterton curve is the least skew and varies from 5 to 91 with only two values over 60.

From Table I it will be seen that the arithmetic mean number of words per sentence is 25·87 for Chesterton, 24·11 for Wells and 31·23 for Shaw. The medians are also different and presumably the quartiles and deciles, but these latter were not calculated.

TABLE I

Frequency constants of the distributions of sentence length

	Chesterton	Wells	Shaw
Number of sentences	600	600	600
Number of words	15,521	14,463	18,735
Arithmetic mean no. of words	25·87	24·11	31·23
Median no. of words	25·3	20·8	26·0
Mean log no. of words	1·37	1·31	1·39
Geometrical mean no. of words	23·5	20·5	24·5
Standard deviation of mean log	0·200	0·237	0·290
Standard error of mean log	0·0080	0·0095	0·0112

If, however, instead of taking the frequency distribution of the actual number of words per sentence we take that of the logarithm of the number we get the distributions shown in Figs. 2–4. They undoubtedly show a very close resemblance to the "normal distribution". The mean log and standard deviation for Chesterton is $1{\cdot}37 \pm 0{\cdot}20$; for Wells $1{\cdot}31 \pm 0{\cdot}24$ and for Shaw $1{\cdot}39 \pm 0{\cdot}29$. The standard error of the mean is, owing to the large number of observations, in all cases very small and approximately $\pm 0{\cdot}01$.

On each of the three figures is superimposed a normal curve of the same area, mean and standard deviation and it will be seen how closely it fits the observed values.

The following comments may, however, be made:

(1) The greater irregularity of the observed values in the lower portion of the distribution is due to the irregular distribution of the logarithms of integers when grouped in small artificial divisions as in the present case. Thus there is no logarithm of an integer between 0·01 and 0·25; none between 0·61 and 0·65 and between 0·71 and 0·75. On the other hand, there are two between 1·11

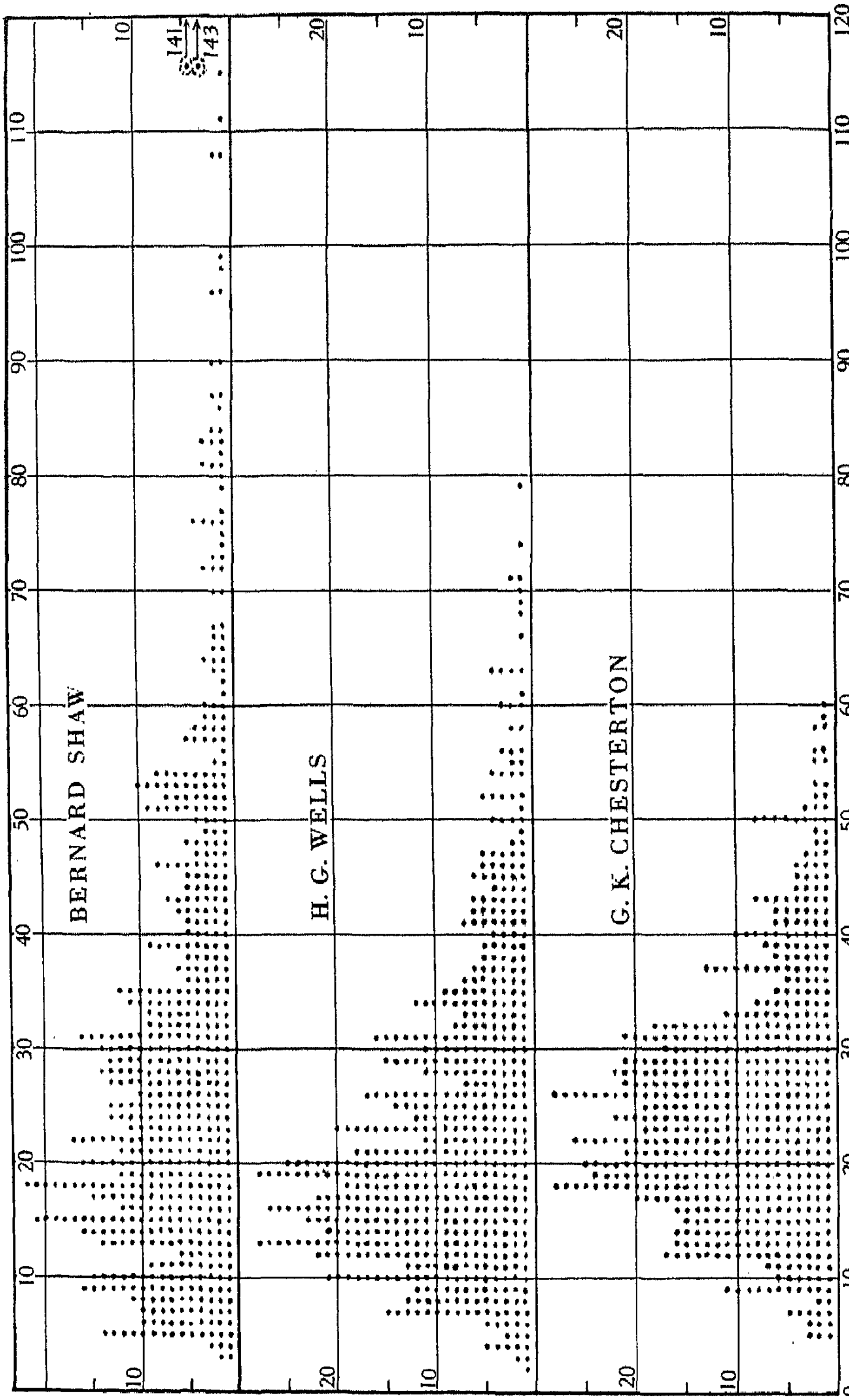

BERNARD SHAW
H. G. WELLS
G. K. CHESTERTON
141
143
0
10
20
30
40
50
60
70
80
90
100
110
120

Fig. 1.

and 1·15; two between 1·16 and 1·20; only one between 1·21 and 1·25 and three between 1·26 and 1·30. Thus in all three diagrams the frequency of 1·21–1·25 is well down and that of 1·26–1·30 is far up. Some process of smoothing would undoubtedly eliminate these irregularities, but it was thought better to leave the data in their original form and draw attention to the sources of irregularity.

(2) There appears to be on all three diagrams a slight shortage of high values and a slight excess of low ones. On the latter I have no comment but it appears possible that the small deficit in the longer sentences might easily be due to a biased effect introduced by the habit that many writers have of cutting up unusually long sentences into their component parts when reading over their manuscript or proofs.

On the assumption that the normal curve is a sound representation of the frequency distribution of the log number of words in sentences, Fig. 5 has been prepared which shows the means and normal curves for the three books superimposed on one another. The means are close together but the distributions are very different.

The difference in means between Shaw and Wells is 0·09 and the standard error of the difference is only 0·015. Thus the difference is six times the standard error and hence certainly significant. Between Shaw and Chesterton the difference of the means is barely significant but that between the standard deviations is quite striking.

If the above reasoning is correct, it is unnecessary for the comparison between two documents to compare arithmetic means, medians, quartiles and deciles, but only the log mean and the log standard deviation; all other comparisons are included in these.

It follows also that Mr Bernard Shaw, while undoubtedly under the impression that he was punctuating at his own free will, was for this particular book hide-bound within the limits of

$$Z = \frac{1}{0{\cdot}29\sqrt{(2\pi)}}\exp\left[\frac{(1{\cdot}4-x)^2}{2(0{\cdot}29)^2}\right],$$

while similarly Mr Wells was writing under the restricting influence of

$$Z = \frac{1}{0{\cdot}24\sqrt{(2\pi)}}\exp\left[\frac{(1{\cdot}3-x)^2}{2(0{\cdot}24)^2}\right],$$

where Z is the frequency and x the logarithm of the number of words per sentence.

It is also perhaps worthy of passing comment that the curve representing Mr Shaw is short and broad while that representing Mr Chesterton is tall and slender; which shows how necessary it is to use these curves only for the purpose for which they were originally designed.

Perhaps something might be added on the meaning in words of the above mathematical transformation. If the log distribution is normal we can infer

that the extent to which a sentence in the process of writing is likely to vary is at any level proportional to the length of the sentence. Thus when he is thinking in short sentences of about 10 words an author is likely to vary say from 8 to 12 words; when he is thinking in longer sentences of say 100 words he will vary from 80 to 120. In other words the variations are proportional or geometric and do not merely involve the addition or subtraction of x words at all levels. Further, if the geometric mean is taken as a basis, sentences between this and half its length are as frequent as those between it and twice its length; sentences down to one-quarter its length are as likely to occur as sentences up to four times its length; and so on.

If the arithmetic mean were the true basis then sentences of 10 words more than the arithmetic mean would be as likely to occur as sentences of 10 words less, and it is easy to see that this is not the case.

Before the whole theory of the use of such distributions for separating works of different authorships can be fully accepted it will of course be necessary to study the results obtained from many different works by the same author, in different styles, on different subjects and at different periods of his life. From these it may be possible to find what variation can occur "within authors" as compared with "between authors". This note is not meant to deal with this basic problem but only to draw attention to the simplification of the method of approach to such a problem by the use of a transformation which produces a normal instead of a skew distribution.

REFERENCES

Williams, C. B. (1927). *Ann. Appl. Ent.* **24**, 404.
Yule, G. Udny (1939). *Biometrika*, **30**, 363–90.

APPLICATIONS OF THE NON-CENTRAL t-DISTRIBUTION

By N. L. JOHNSON and B. L. WELCH

1. Introduction

Statistical problems arising in connexion with the normal distribution are simplified by the circumstance that the sample mean $\bar{x}$ and standard deviation s* are jointly sufficient estimates of the corresponding population parameters ξ and σ. Also the distributions of $\bar{x}$ and s have simple forms and together with Student's t-distribution provide complete solutions of any problems of testing hypotheses or of estimating fiducial limits relating to either ξ or σ singly. In the present paper we shall consider some of the questions which arise when our main concern is not with either ξ or σ alone, but with some function of the two. In particular we shall consider a number of cases which all lead to the use of what has been called the non-central t-distribution. Tables of the probability integral of this distribution will be given in a form suitable for the solution of the problems raised.

2. The non-central t-distribution

Let z be a quantity distributed normally about zero with unit standard deviation and let w be a quantity distributed independently as χ^2/f, where f is the number of degrees of freedom of the χ^2. Then if t is defined by the equation

$$t = \frac{z+\delta}{\sqrt{w}}, \qquad \ldots\ldots(1)$$

where δ is some constant, then t is distributed in a manner depending only on δ and f. This distribution will be termed a non-central t-distribution. When δ equals zero we have the familiar Student's t.

In general the elementary probability distribution of t is given by

$$p(t) = \frac{1}{2^{\frac{1}{2}(f-1)}\Gamma(\frac{1}{2}f)\sqrt{\pi}} e^{-\frac{1}{2}\frac{f\delta^2}{f+t^2}} \left(\frac{f}{f+t^2}\right)^{\frac{1}{2}(f+1)} \int_0^\infty v^f e^{-\frac{1}{2}\left(v - \frac{t\delta}{\sqrt{(f+t^2)}}\right)^2} dv$$

$$= \frac{f!}{2^{\frac{1}{2}(f-1)}\Gamma(\frac{1}{2}f)\sqrt{\pi}} e^{-\frac{1}{2}\frac{f\delta^2}{f+t^2}} \left(\frac{f}{f+t^2}\right)^{\frac{1}{2}(f+1)} Hh_f\left(-\frac{t\delta}{\sqrt{(f+t^2)}}\right), \qquad \ldots\ldots(2)$$

where

$$Hh_f(x) = \int_0^\infty \frac{v^f}{f!} e^{\frac{1}{2}(v+x)^2} dv. \qquad \ldots\ldots(3)$$

An account of some of the properties of the above-defined Hh function has been given by R. A. Fisher (1931). He derives a result equivalent to the above equation (2) although with a different notation, his t being equivalent to our $t/\sqrt{f}$ and his

* $s^2 = \Sigma(x-\bar{x})^2/(n-1)$, where n is the sample size.

τ being our $\delta/\sqrt{(f+1)}$. Tables of the Hh function have been calculated by J. R. Airey (1931). These tables, however, are not useful for solving the problems which are considered in the present paper. For these problems the probability integral of the non-central t-distribution is required. Existent tables of this integral (J. Neyman, 1935 and J. Neyman & B. Tokarska, 1936) have been calculated only for one rather restricted purpose. In the present paper much more extensive tables will be provided which, it is hoped, will cover all the applications of the non-central t-distribution likely to be encountered.

The probability integral of non-central t demands a table of triple entry, since the probability that t exceeds t_0, say, depends on f, δ, and t_0. The notations

$$P(f,\delta,t_0) = P(t>t_0\,|\,f,\delta) \qquad \ldots\ldots(4)$$

will be used to denote this probability.

Often it is necessary to find what value of t_0 is such that the probability (4) will take a specified value ϵ, say. This t_0 will be a function of f, δ, and ϵ and it will be convenient to denote it by $t(f,\delta,\epsilon)$. Thus

$$P\{t>t(f,\delta,\epsilon)\,|\,f,\delta\} = \epsilon.$$

Again, often t_0 will be given and the value of δ which makes (4) take the value ϵ, will be required. It will be convenient to denote this value of δ by $\delta(f,t_0,\epsilon)$. Thus

$$P\{t>t_0\,|\,f,\delta(f,t_0,\epsilon)\} = \epsilon.$$

Space does not permit the tabling of all the three functions $P(f,\delta,t_0)$, $t(f,\delta,\epsilon)$ and $\delta(f,t_0,\epsilon)$. For reasons which will become clear later in the paper it was decided to table $\delta(f,t_0,\epsilon)$ most fully. Table IV, given at the end of the paper, facilitates the direct calculation of $\delta(f,t_0,\epsilon)$ for seventeen probability levels ϵ. It can also be used without difficulty for calculating $t(f,\delta,\epsilon)$. Table V is an additional short table from which $t(f,\delta,\epsilon)$ can be calculated rather more directly but only for the probability levels $\epsilon=0{\cdot}05$ and $\epsilon=0{\cdot}95$. These tables are given in the most suitable form which we have been able to evolve consistent with necessary limitations on size.

Before the tables are described, the next three sections will be devoted to a description of some of the situations where they are required.

3. COEFFICIENT OF VARIATION

The first function of ξ and σ to be discussed will be the coefficient of variation $V=\sigma/\xi$. In the practical situations where this index is an appropriate measure of variability, the variable x is usually necessarily positive. Now for a normal population the ratio of the mean to the standard deviation has to be of the order of 3 or more, for the chance of a negative x to be negligible. Strictly speaking, therefore, the sampled population should not be assumed normal if the coefficient of variation is too great. The figure 33 % has often been stated as the permissible

upper limit. In practice coefficients of variation are usually much smaller than this, and it is assumed in the present section that we are dealing with such cases.

An estimate of V is provided by the sample coefficient of variation $v = s/\bar{x}$. Now since we may write

$$\frac{\sqrt{n}}{v} = \frac{\sqrt{n}\,\bar{x}}{s} = \left\{\frac{\sqrt{n}(\bar{x}-\xi)}{\sigma} + \frac{\sqrt{n}\,\xi}{\sigma}\right\} \div \frac{s}{\sigma} \qquad \text{......(5)}$$

it appears from comparison with (1) that $\sqrt{n}/v$ is distributed as non-central t with $f = (n-1)$ and $\delta = \sqrt{n}/V$ (McKay, 1932). The solution of problems relating to V is therefore easily effected.

For instance, suppose it is desired to test whether the sample contradicts the hypothesis that $V \leqslant V_0$, and that we decide to reject the hypothesis when $v > v_0$, where v_0 is chosen so that $P(v > v_0 \mid V = V_0)$ is equal to some specified small chance ϵ. In the notation of the previous section v_0 will then be given by

$$\frac{\sqrt{n}}{v_0} = t\left(\overline{n-1}, \frac{\sqrt{n}}{V_0}, \overline{1-\epsilon}\right). \qquad \text{......(6)}$$

Again, consider an example where it is decided to reject a sample as unsatisfactory when v is greater than a given value v_0, and where it is required to know how low the true coefficient of variability should be kept to ensure that the probability of rejection will not exceed a given ϵ, i.e. we require V_0 such that $P(v > v_0 \mid V = V_0) = \epsilon$. In the notation of Section 2, V_0 is given by

$$\frac{\sqrt{n}}{V_0} = \delta\left(\overline{n-1}, \frac{\sqrt{n}}{v_0}, \overline{1-\epsilon}\right). \qquad \text{......(7)}$$

Or finally suppose that a value of v is observed and an upper fiducial limit of V is required so that the chance is ϵ of this limit being exceeded, i.e. a lower fiducial limit of $\sqrt{n}/V$ is required. Now since

$$P\left\{\frac{\sqrt{n}}{v} > t\left(\overline{n-1}, \frac{\sqrt{n}}{V}, \epsilon\right)\right\} = \epsilon$$

and since the inequality $\quad \dfrac{\sqrt{n}}{v} > t\left(\overline{n-1}, \dfrac{\sqrt{n}}{V}, \epsilon\right)$

is equivalent to the inequality*

$$\frac{\sqrt{n}}{V} < \delta\left(\overline{n-1}, \frac{\sqrt{n}}{v}, \epsilon\right),$$

it is seen that the required upper fiducial limit of V is

$$\sqrt{n}\Big/\delta\left(\overline{n-1}, \frac{\sqrt{n}}{v}, \epsilon\right). \qquad \text{......(8)}$$

* This follows from the fact that $t(f, \delta, \epsilon)$ is a monotonically increasing function of δ.

4. THE POWER OF STUDENT'S t-TEST

Suppose it is desired to test the hypothesis H_0 that the mean ξ of a normal population has the value ξ_0. Student's t-test consists in calculating from the data the quantity

$$t = \frac{\sqrt{n}\ (\bar{x}-\xi_0)}{s} \qquad \text{......(9)}$$

and referring it to the usual central t-distribution with $f = (n-1)$ degrees of freedom. Thus if the only alternative hypotheses to be taken into consideration are those for which $\xi > \xi_0$, the test will consist in rejecting the hypothesis when $t > t_0$, where t_0 is such that $P(t > t_0 \mid H_0) = \epsilon$, and ϵ is the conventional level of significance. In our notation

$$t_0 = t(\overline{n-1},\ 0,\ \epsilon). \qquad \text{......(10)}$$

Now when H_0 is not true, but ξ has some alternative value ξ_1, it is often required to know how powerful Student's test will be, i.e. what chance the test will have of rejecting H_0. But we can write (9) in the form

$$t = \left\{\frac{\sqrt{n}\ (\bar{x}-\xi_1)}{\sigma} + \frac{\sqrt{n}\ (\xi_1-\xi_0)}{\sigma}\right\} \div \frac{s}{\sigma} \qquad \text{......(11)}$$

whence, comparing with (1), it now appears that the quantity calculated from the data is distributed in the non-central t-distribution with $f = (n-1)$ and

$$\delta = \sqrt{n}\ (\xi_1-\xi_0)/\sigma.$$

The power of the test is therefore $P(\overline{n-1}, \sqrt{n}\ (\xi_1-\xi_0)/\sigma, t_0)$ and the value of ξ_1 for which the power reaches any specified value η, say, is given by

$$\frac{\sqrt{n}\ (\xi_1-\xi_0)}{\sigma} = \delta(\overline{n-1}, t_0, \eta). \qquad \text{......(12)}$$

Tables for evaluating the power of the t-test together with a discussion of their use have been given by J. Neyman (1935) and J. Neyman & B. Tokarska (1936). They are not restricted to the simple case of testing whether a mean has a specified value but apply to all cases in which the t-test is used.

5. PROPORTION DEFECTIVE PROBLEMS

Another class of problem where the non-central t-distribution has an application, occurs when objects are classified as defective or non-defective according to whether they have values of a characteristic exceeding or falling short of a fixed standard. Thus, if an object is defective when the character x exceeds a fixed given level L, then information will often be required about the proportion P falling beyond L in the population. If the population be normal P clearly depends only on the ratio

$$U = \frac{(L-\xi)}{\sigma}.$$

An estimate of U is provided from the sample by calculating

$$u = \frac{(L-\bar{x})}{s}. \qquad \text{......(13)}$$

Corresponding to deviate u the equivalent normal probability p will then give an estimate of P. This situation has been discussed recently by W. J. Jennett & B. L. Welch (1939).* The transference from U to P and from u to p does of course require that the sampled distribution is really approximately normal and for this reason care must be taken in going out to the "tails" of the distribution (i.e. when U is large and P small).

In order to allow for the sampling errors arising from this method of estimating P we have only to note that (13) may be written

$$\sqrt{n}\, u = \left\{\frac{\sqrt{n}\,(L-\xi)}{\sigma} - \frac{\sqrt{n}\,(\bar{x}-\xi)}{\sigma}\right\} \div \frac{s}{\sigma}. \qquad \text{......(14)}$$

Hence $\sqrt{n}\, u$ is distributed as non-central t with $f = (n-1)$ and $\delta = \sqrt{n}\, U$. Thus if the proportion P (and hence U) were known the value u_0 of u such that $P(u > u_0) = \epsilon$ would be given by

$$\sqrt{n}\, u_0 = t(\overline{n-1}, \sqrt{n}\, U, \epsilon). \qquad \text{......(15)}$$

Conversely, given u from a sample, a lower fiducial limit of U is obtained by noting that the inequality

$$\sqrt{n}\, u > t(\overline{n-1}, \sqrt{n}\, U, \epsilon)$$

is equivalent to the inequality

$$\sqrt{n}\, U < \delta(\overline{n-1}, \sqrt{n}\, u, \epsilon). \qquad \text{......(16)}$$

Hence a lower fiducial limit for U will be given by $\delta(\overline{n-1}, \sqrt{n}\, u, \epsilon)/\sqrt{n}$. In the above analysis the proportion falling beyond L has been termed a "proportion defective". In industrial problems this is often a fair description. More generally the analysis refers to any problem where we are primarily interested in how a measurable object is related to a fixed level. A slightly different problem occurs when we are not given L but are asked to estimate the value of L so that P will have an assigned value. For instance, the value of L may be required so that the chance is 1 in 20 of it being exceeded.

Now if U is the normal deviate which is exceeded with probability P, then the required L is related to ξ and σ by the relation

$$L = \xi + U\sigma.$$

An estimate of L is therefore given by

$$l = \bar{x} + Us. \qquad \text{......(17)}$$

* For a general discussion of similar problems arising in the control of industrial products the reader may be referred to the British Standards Institution Publication, No. 600 (E. S. Pearson, 1935).

If fiducial limits are required for L they may be obtained by returning to equations (13) and (15). L is not now given, but U is definitely known and therefore (15) provides the value of u_0 such that $P(u > u_0) = \epsilon$. But the inequality

$$u = \frac{L - \bar{x}}{s} > u_0$$

can be reversed to read $$L > \bar{x} + u_0 s.$$

Hence, if we write $$l_\epsilon = \bar{x} + \frac{s t(\overline{n-1, \sqrt{n}\, U, \epsilon})}{\sqrt{n}} \qquad \text{......(18)}$$

we shall have $P(L > l_\epsilon) = \epsilon$. Hence l_ϵ will be a fiducial limit for L, which will be exceeded by L in a proportion ϵ of cases.

6. Large sample results

Before going on to discuss the distribution of non-central t in general and the method of applying the tables given below, it will be convenient to consider the situation when f is large. The distribution then approaches the normal form. The important question is how quickly.

The first three moments of t are given exactly by the following expressions:

$$\mu_1' = \sqrt{\frac{f}{2}}\, \frac{\Gamma\left(\frac{f}{2} - \frac{1}{2}\right)}{\Gamma\left(\frac{f}{2}\right)}\, \delta, \qquad \text{......(19)}$$

$$\mu_2 = \frac{f(1+\delta^2)}{(f-2)} - \mu_1'^2, \qquad \text{......(20)}$$

$$\mu_3 = \mu_1' \left[\frac{f(2f - 3 + \delta^2)}{(f-2)(f-3)} - 2\mu_2\right]. \qquad \text{......(21)}$$

If the gamma functions are expanded in powers of $1/f$ and only the leading terms are retained, these give *approximately*

$$\mu_1' = \delta; \quad \mu_2 = \left(1 + \frac{\delta^2}{2f}\right); \quad \mu_3 = \frac{\delta}{f}\left[3 + \frac{5\delta^2}{4f}\right]; \quad \sqrt{\beta_1} = \frac{\delta\left[3 + \frac{5\delta^2}{4f}\right]}{f\left[1 + \frac{\delta^2}{2f}\right]^{\frac{3}{2}}}. \qquad \text{......(22)*}$$

In large samples, therefore, t becomes normally distributed about δ with standard deviation $\left(1 + \frac{\delta^2}{2f}\right)^{\frac{1}{2}}$. The rapidity of the approach to symmetry is indicated by the true values of $\sqrt{\beta_1}$ in Table I. For given f the greatest value $\sqrt{\beta_1}$ can take is shown in the last column.

* The term δ^2/f is retained because in most problems δ will be of the order $\sqrt{f}$.

TABLE I

$\sqrt{\beta_1}$ *of t for different f and* δ

f \ $\delta/\sqrt{f}$	0	1	2	3	∞
4	0·00	3·70	4·61	4·86	5·09
6	0·00	1·67	2·13	2·26	2·38
8	0·00	1·20	1·55	1·65	1·74
12	0·00	0·84	1·10	1·18	1·25
24	0·00	0·53	0·69	0·74	0·79
∞	0·00	0·00	0·00	0·00	0·00

It is seen from Table I that values of $\sqrt{\beta_1}$ of the order of unity are likely to occur fairly frequently in the type of problem which has been discussed in the previous sections. This represents a considerable degree of skewness and another method of making normal approximations, outlined below, is preferable. Before describing this we may note that if t is referred to a normal scale with mean δ and standard deviation $\left(1+\frac{\delta^2}{2f}\right)^{\frac{1}{2}}$ then the following approximations to the quantities defined in section 2 will be obtained:

$$t(f,\delta,\epsilon) = \delta+K_\epsilon\left(1+\frac{\delta^2}{2f}\right)^{\frac{1}{2}}, \qquad \text{......(23)}$$

where K_ϵ is the deviate of the unit normal curve exceeded with a probability ϵ. Also $\delta(f, t_0, \epsilon)$ will be given by the solution of the equation in δ

$$t_0 = \delta+K_\epsilon\left(1+\frac{\delta^2}{2f}\right)^{\frac{1}{2}}. \qquad \text{......(24)}$$

On rationalization this equation is quadratic in δ and the correct root to take will be obvious.

Better approximations, which are, however, still based on the normal curve may be obtained as follows. The probability that t exceeds a given value t_0 is by (1) the probability of the inequality

$$t = \frac{(z+\delta)}{\sqrt{w}} > t_0, \qquad \text{......(25)}$$

and this inequality is equivalent to

$$(-z+t_0\sqrt{w}) < \delta. \qquad \text{......(26)}$$

Now $(-z)$ is a unit normal deviate, and $\sqrt{w}$, being of the form $\chi/\sqrt{f}$ is very nearly normally distributed even for small f. Since z and $\sqrt{w}$ are independent, $(-z+t_0\sqrt{w})$ must therefore be more nearly normally distributed than $\sqrt{w}$, whatever t_0.*

* This is practically obvious, but a demonstration is given in the Appendix, equation (43).

Hence an upper limit to the skewness of $(-z+t_0\sqrt{w})$ is given for different f by the values of $\sqrt{\beta_1}$ in Table II. Comparison with Table I shows that it is better to take $(-z+t_0\sqrt{w})$ as normally distributed rather than t. The procedure is then as follows:

Write
$$\text{mean}\sqrt{w} = a; \quad \sigma_{\sqrt{w}} = \frac{b}{\sqrt{(2f)}}. \qquad \ldots\ldots(27)$$

Then $(-z+t_0\sqrt{w})$ is taken to be normally distributed with mean at_0 and standard deviation $\left(1+\frac{b^2t_0^2}{2f}\right)^{\frac{1}{2}}$. For given t_0 the value of δ such that $P(t>t_0)$ equals ϵ is given by seeking δ such that the probability is ϵ of (26) being true. This gives

$$\delta(f, t_0, \epsilon) = at_0 - K_\epsilon\left(1+\frac{b^2t_0^2}{2f}\right)^{\frac{1}{2}}. \qquad \ldots\ldots(28)$$

TABLE II

$\sqrt{\beta_1}$ of $\sqrt{w}$ for different f

f	4	6	8	12	24	∞
$\sqrt{\beta_1}$	0·41	0·32	0·27	0·21	0·15	0·00

Conversely for given δ the value $t(f, \delta, \epsilon)$ which will be exceeded with probability ϵ, will be approximated by solving for t the equation

$$\delta = at - K_\epsilon\left(1+\frac{b^2t^2}{2f}\right)^{\frac{1}{2}}. \qquad \ldots\ldots(29)$$

On rationalization this becomes a quadratic in t. This method of approximation was given by W. J. Jennett & B. L. Welch (1939) together with a short table of a and b. However, since a and b differ from unity by a quantity of the order $1/f$ and since errors at least of this magnitude are involved in the assumption of normality, it would perhaps have been as logical to take a and b equal to unity in the approximation. The numerical comparisons set out in Table III indicate that nothing is lost by doing this and also show how inferior is the method of approximation which assumes t itself to be normally distributed.

Therefore the approximation which we shall adopt as being the best one, requiring only the use of the normal probability scale, will consist in assuming $(-z+t_0\sqrt{w})$ to be normally distributed about t_0 with standard deviation

$$\left(1+\frac{t_0^2}{2f}\right)^{\frac{1}{2}}.$$

The value of δ such that $p(t>t_0)$ equals ϵ will then be approximated by

$$\delta(f, t_0, \epsilon) = t_0 - K_\epsilon\left(1+\frac{t_0^2}{2f}\right)^{\frac{1}{2}}. \qquad \ldots\ldots(30)$$

TABLE III

Values of $\delta/\sqrt{f}$

f	ϵ	$\frac{t_0}{\sqrt{f}}$	$\delta/\sqrt{f}$			
			True	Approx. 1	Approx. 2	Approx. 3
4	0·99	2	0·056	0·736	−0·088	−0·015
4	0·01	2	3·972	12·023	3·848	4·015
4	0·99	5	1·120	2·668	0·565	0·726
4	0·01	5	9·247	29·230	8·840	9·274
24	0·99	2	1·181	1·345	0·956	1·178
24	0·01	2	2·819	3·163	2·800	2·822
24	0·99	5	3·290	3·677	3·213	3·255
24	0·01	5	6·752	7·593	6·687	6·745

The values correspond to certain f, ϵ and $t_0/\sqrt{f}$ and have the property that $p(t>t_0 \mid f, \delta) = \epsilon$. Approximation 1 is the value obtained from equation (23), approximation 2 by using (28) with the correct a and b and approximation 3 by using (30), i.e. taking $a = b = 1$.

Conversely, for given δ, the value $t(f, \delta, \epsilon)$ which will be exceeded with probability ϵ will be given implicitly by

$$\delta = t - K_\epsilon\left(1+\frac{t^2}{2f}\right)^{\frac{1}{2}}, \qquad \text{......(31)}$$

which on rationalizing and solving becomes

$$t(f, \delta, \epsilon) = \frac{\delta + K_\epsilon\left(1+\dfrac{\delta^2}{2f}-\dfrac{K_\epsilon^2}{2f}\right)^{\frac{1}{2}}}{\left(1-\dfrac{K_\epsilon^2}{2f}\right)}. \qquad \text{......(32)}$$

The approximations (30) and (32) are useful even in moderate-sized samples. It is not intended, however, to enter here into any detailed discussion of how good the various approximations are. They have been considered primarily because they form the basis of the exact tables which are given at the end of the paper and which are described in the next section.

7. Description and use of tables

The tables are intended to facilitate the exact calculation of the quantities $\delta(f, t_0, \epsilon)$ and $t(f, \delta, \epsilon)$ defined in Section 2. The form in which they are presented was determined largely by the fact that the space they had to occupy was limited, and also to some extent by the method which was adopted to calculate them.

Table IV may be used to give $\delta(f, t_0, \epsilon)$ for 17 probability levels, viz.

$$\epsilon = 0{\cdot}005, 0{\cdot}01, 0{\cdot}025, 0{\cdot}05, 0{\cdot}1, 0{\cdot}2, 0{\cdot}3, 0{\cdot}4, 0{\cdot}5, 0{\cdot}6, 0{\cdot}7,$$
$$0{\cdot}8, 0{\cdot}9, 0{\cdot}95, 0{\cdot}975, 0{\cdot}99 \text{ and } 0{\cdot}995.$$

The basis of the table is the large sample approximation described in the last section, equation (30). The K_ϵ in this equation is simply the normal multiple which is exceeded with probability ϵ. In small samples an error will be committed in taking this form of approximation, and in Table IV we give instead a multiple $\lambda(f, t_0, \epsilon)$, which is such that *exactly*

$$\delta(f, t_0, \epsilon) = t_0 - \lambda(f, t_0, \epsilon)\left(1 + \frac{t_0^2}{2f}\right)^{\frac{1}{2}}. \qquad \ldots\ldots(33)$$

Even for samples giving f as small as 4, it is found that λ never differs much from K_ϵ. This stability makes interpolation easy so that, for given ϵ, it is possible to provide a table of small compass covering all values of t_0 from $-\infty$ to ∞.

For given ϵ the tabled quantity λ is a function of f and t_0 and accordingly Table IV, for given ϵ, is one of double entry.

Values are given corresponding to $f = 4, 5, 6, 7, 8, 9, 16, 36, 144$ and ∞. The reason for choosing these values of f is that λ differs from K_ϵ by a quantity of the order $1/\sqrt{f}$. The values $f = 9, 16, 36, 144$ and ∞ are such that $12/\sqrt{f} = 4, 3, 2, 1$ and 0 respectively. Hence interpolations for intermediate f's may be simply effected by considering λ as a function of $12/\sqrt{f}$, since we are then dealing with a function tabled at equal intervals.

The choice of the values of t_0 in the table is also determined by the necessity for interpolation to be simple and yet the table not to be unduly large. The whole range of t_0 from $-\infty$ to ∞ has to be covered. This has been done as follows. For $t_0/\sqrt{2f}$ between $-\infty$ and $-0{\cdot}75$, λ is given against the quantity

$$y = \left(1 + \frac{t_0^2}{2f}\right)^{-\frac{1}{2}}. \qquad \ldots\ldots(34)$$

For $t_0/\sqrt{2f}$ between $-0{\cdot}75$ and $0{\cdot}75$, λ is tabled against

$$y' = \frac{t_0}{\sqrt{2f}}\left(1 + \frac{t_0^2}{2f}\right)^{-\frac{1}{2}} \qquad \ldots\ldots(35)$$

For $t_0/\sqrt{2f}$ between $0{\cdot}75$ and ∞, λ is again tabled against y. The argument interval for both y and y' is $0{\cdot}1$. It will be noted that $y = (1 - y'^2)^{\frac{1}{2}}$.

From the point of view of interpolation other, perhaps simpler, functions of t_0 could have been used in constructing the table. The choice of y was made because the quantity $\left(1 + \frac{t_0^2}{2f}\right)^{\frac{1}{2}}$ is wanted in any case for substitution into equation (33) after λ has been obtained. y' had to be used in addition because of interpolation difficulties in the middle of the table.

It will be noted that in Table IV only the double entry tables for $\epsilon = 0{\cdot}005, 0{\cdot}01, 0{\cdot}025, 0{\cdot}05, 0{\cdot}1, 0{\cdot}2, 0{\cdot}3, 0{\cdot}4$ and $0{\cdot}5$ are given. For $\epsilon > 0{\cdot}5$ we can use the fact that

$$\delta(f, t_0, \epsilon) = -\delta(f, -t_0, \overline{1-\epsilon}), \qquad \ldots\ldots(36)$$

a relation which is apparent from the form of equation (1).

To summarize, the steps necessary to find the value of δ such that $P(t > t_0) = \epsilon$ are as follows.

(i) Find $y = \left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}}$ or $y' = \frac{t_0}{\sqrt{2f}}\left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}}$ according as $\left|\frac{t_0}{\sqrt{2f}}\right|$ is greater than or less than 0·75.

(ii) If $f > 9$ calculate $12/\sqrt{f}$.

(iii) If ϵ is one of the values 0·005, ..., 0·5, enter the appropriate part of Table IV and obtain $\lambda(f, t_0, \epsilon)$, by interpolating with respect to the quantities obtained in (i) and (ii).

(iv) Calculate $$\delta(f, t_0, \epsilon) = t_0 - \lambda\left(1+\frac{t_0^2}{2f}\right)^{\frac{1}{2}}.$$

(v) If ϵ is one of the values 0·6, 0·7, ..., 0·995 calculate $\delta(f, -t_0, \overline{1-\epsilon})$ and then change its sign.

Inverse use of Table IV

Next consider the calculation of $t(f, \delta, \epsilon)$, i.e. the situation where δ is given and t_0 is required so that $P(t > t_0) = \epsilon$. In the previous section it was shown that a first approximation will be given by equation (32). The true relation between $t(f, \delta, \epsilon)$ and δ will, however, be obtained by replacing K_ϵ in (32) by the λ which is tabled in Table IV; thus

$$t(f, \delta, \epsilon) = \frac{\delta + \lambda\left(1+\frac{\delta^2}{2f}-\frac{\lambda^2}{2f}\right)^{\frac{1}{2}}}{\left(1-\frac{\lambda^2}{2f}\right)}. \qquad \text{......(37)}$$

The drawback about this equation is that λ is given in Table IV as a function of t_0, and t_0 is not now known. An iteration method is therefore necessary.

The successive steps in this iteration may be summarized as follows:

(i) The first approximation t_1 is given by (32); thus

$$t_1 = \frac{\delta + K_\epsilon\left(1+\frac{\delta^2}{2f}-\frac{K_\epsilon^2}{2f}\right)^{\frac{1}{2}}}{\left(1-\frac{K_\epsilon^2}{2f}\right)}.$$

(ii) Next use Table IV to calculate directly

$$\lambda_1 = \lambda(f, t_1, \epsilon).$$

(iii) A second approximation t_2 is then given by substituting this λ_1 in (37); thus

$$t_2 = \frac{\delta + \lambda_1\left(1+\frac{\delta^2}{2f}-\frac{\lambda_1^2}{2f}\right)^{\frac{1}{2}}}{\left(1-\frac{\lambda_1^2}{2f}\right)}.$$

(iv) Then by finding $\lambda_2 = \lambda(f, t_2, \epsilon)$ and substituting in (37) a third approximation t_3 will be obtained. These steps must be repeated until two successive approximations give the same value. In practice this is found to occur very quickly so that there is likely to be no need of more than three approximations.

(v) If ϵ is one of the values 0·6, 0·7, ..., 0·995 calculate $t(f, -\delta, \overline{1-\epsilon})$ and then change its sign.

Use of Table V

The iteration process described above is not very lengthy in actual practice, but even this amount of trouble would be unnecessary if λ were tabled as a function of δ as well as a function of t. Such a table, similar in extent to Table IV, could of course be calculated, but it did not seem worth while to do this, since the inverse method of using Table IV is not difficult. However, it has been thought useful to give for the single probability level $\epsilon = 0{\cdot}05$, a table which can be entered with δ, since very often this conventional 1 in 20 chance is taken as a point of reference in statistical problems.

In Table V, λ is tabled against

$$\eta' = \frac{\delta}{\sqrt{(2f)}}\left(1+\frac{\delta^2}{2f}\right)^{-\frac{1}{2}} \qquad \text{......(38)}$$

at intervals of 0·1 from -1 to 1. When δ is given, this function of δ may be quickly calculated and then λ may be obtained by interpolating in Table V. The substitution of this value of λ in (37) gives the required $t(f, \delta, 0{\cdot}05)$. As before $t(f, \delta, 0{\cdot}95)$ may also be calculated from the same table, being equal to *minus* $t(f, -\delta, 0{\cdot}05)$.

8. EXAMPLES

In the present section some numerical examples will be worked to illustrate the application of Tables IV and V. Reference will be made to theoretical results obtained in Sections 3–5.

Example 1. In a sample of $n = 25$ a coefficient of variation $v = 2{\cdot}6$ is observed. Obtain from this a "median" estimate of the population coefficient V. By a "median" estimate is meant one for which the chance of it exceeding the true V is 0·50. In other words our estimate is the 50 % fiducial limit. Hence from (8) the required estimate is

$$v_m = \sqrt{25}/\delta(f, t_0, 0{\cdot}5), \qquad \text{......(39)}$$

where $$t_0 = \frac{\sqrt{25}}{2{\cdot}6} = 1{\cdot}9231 \text{ and } f = 24.$$

We have $$\left(1+\frac{t_0^2}{2f}\right)^{\frac{1}{2}} = 1{\cdot}0379; \quad y' = \frac{t_0}{\sqrt{(2f)}}\left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}} = 0{\cdot}2674,$$

$$12/\sqrt{f} = 2{\cdot}4495.$$

From Table IV, for $\epsilon = 0{\cdot}5$, we obtain by linear interpolation* with respect to y' and $12/\sqrt{f}$, $\lambda = 0{\cdot}0197$. Hence from (33)

$$\delta = 1{\cdot}9231 - (0{\cdot}0197)(1{\cdot}0379) = 1{\cdot}9027.$$

Therefore from (39), $v_m = 2{\cdot}628$.

Example 2. A sample of $n = 10$ measurements of a normally distributed character x is given. The mean $\bar{x}$ is 4·7 and the standard deviation s is 0·2. Obtain 10 % fiducial limits for the proportion P in the population exceeding $x = 5{\cdot}0$.

As is shown in Section 5 this is equivalent to finding limits for $U = (5{\cdot}0 - \xi)/\sigma$. We calculate first $u = (5{\cdot}0 - \bar{x})/s = 1{\cdot}5$. Substituting this into (16) the upper and lower 10 % points for U are given by $\delta(9, 4{\cdot}743, 0{\cdot}9)/\sqrt{10}$ and $\delta(9, 4{\cdot}743, 0{\cdot}1)/\sqrt{10}$.

We have
$$\left(1+\frac{t_0^2}{2f}\right)^{\frac{1}{2}} = 1{\cdot}5; \quad y = \left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}} = 0{\cdot}6667.$$

From Table IV for $\epsilon = 0{\cdot}1$ we obtain by linear interpolation*

$$\lambda(9, 4{\cdot}743, 0{\cdot}1) = 1{\cdot}347,$$

and hence from (33)

$$\delta(9, 4{\cdot}743, 0{\cdot}1) = 4{\cdot}743 - (1{\cdot}347)(0{\cdot}6667) = 2{\cdot}723.$$

To obtain $\delta(9, 4{\cdot}743, 0{\cdot}9)$ we note by (36) that this is equivalent to *minus* $\delta(9, -4{\cdot}743, 0{\cdot}1)$. The corresponding λ is obtained by entering the same table as before, with the same value of y, but in the part of the table for which t_0 is negative. We obtain

$$\lambda(9, -4{\cdot}743, 0{\cdot}1) = 1{\cdot}197.$$

Hence
$$\delta(9, -4{\cdot}743, 0{\cdot}1) = -4{\cdot}743 - (1{\cdot}197)(0{\cdot}6667) = -6{\cdot}538,$$

and therefore $\delta(9, 4{\cdot}743, 0{\cdot}9) = 6{\cdot}538$.

The upper and lower 10 % limits for U are therefore $6{\cdot}538/\sqrt{10}$ and $2{\cdot}723/\sqrt{10}$, i.e. 2·068 and 0·861. The corresponding limits for P are 0·019 and 0·195.

Example 3. Suppose it is required to estimate fiducial limits for the value L of x in the normal population which is such that the proportion of the population exceeding L is 10 %. Consider the case where the sample size is $n = 10$ and the limits required are the 5 % limits at each end. From (18), we require

$$\bar{x} + k_{0{\cdot}05}\, s \quad \text{and} \quad \bar{x} + k_{0{\cdot}95}\, s,$$

where
$$k_\epsilon = \frac{t(9, \sqrt{10}\, U, \epsilon)}{\sqrt{10}}, \qquad \ldots\ldots(40)$$

and U is the normal deviate exceeded with probability 0·1; i.e. $U = 1{\cdot}2816$. We therefore require $t(9, 4{\cdot}053, 0{\cdot}05)$ and $t(9, 4{\cdot}053, 0{\cdot}95)$. In the first place these will be derived from Table IV by the inverse method described above.

* For a note on the accuracy of linear interpolation see the end of Section 9.

Following out the successive steps to derive $t(9, 4\cdot053, 0\cdot05)$ we obtain

$$\text{(i)}\ t_1 = \frac{4\cdot053 + (1\cdot6449)(1 + 0\cdot9126 - 0\cdot1503)^{\frac{1}{2}}}{(1 - 0\cdot1503)} = 7\cdot340,$$

$$\text{(ii)}\ y_1 = \left(1 + \frac{t_1^2}{2f}\right)^{-\frac{1}{2}} = 0\cdot5004; \quad \lambda_1 = \lambda(9,\ 7\cdot340,\ 0\cdot05) = 1\cdot6803,$$

$$\text{(iii)}\ t_2 = \frac{4\cdot053 + (1\cdot6803)(1 + 0\cdot9126 - 0\cdot1569)^{\frac{1}{2}}}{(1 - 0\cdot1569)} = 7\cdot448,$$

$$\text{(iv)}\ y_2 = \left(1 + \frac{t_2^2}{2f}\right)^{-\frac{1}{2}} = 0\cdot4950; \quad \lambda_2 = \lambda(9,\ 7\cdot448,\ 0\cdot05) = 1\cdot6800,$$

$$t_3 = 7\cdot447.$$

To obtain $t(9, 4\cdot053, 0\cdot95)$ we require first $t(9, -4\cdot053, 0\cdot05)$. The successive approximations to this quantity are

$$\text{(i)}\ t_1 = \frac{-4\cdot053 + (1\cdot6449)(1 + 0\cdot9126 - 0\cdot1503)^{\frac{1}{2}}}{(1 - 0\cdot1503)} = -2\cdot200,$$

$$\text{(ii)}\ y_1' = \frac{t_1}{\sqrt{(2f)}}\left(1 + \frac{t_1^2}{2f}\right)^{-\frac{1}{2}} = -0\cdot4602; \quad \lambda_1 = \lambda(9,\ -2\cdot200,\ 0\cdot05) = 1\cdot5933,$$

$$\text{(iii)}\ t_2 = -2\cdot249,$$

$$\text{(iv)}\ y_2' = \frac{t_2}{\sqrt{(2f)}}\left(1 + \frac{t_2^2}{2f}\right)^{-\frac{1}{2}} = -0\cdot4684; \quad \lambda_2 = \lambda(9,\ -2\cdot249,\ 0\cdot05) = 1\cdot5925,$$

$$t_3 = -2\cdot250;$$

$t(9, 4\cdot053, 0\cdot95)$ is equal to *minus* $t(9, -4\cdot053, 0\cdot05)$ and therefore equals $2\cdot250$.

Hence from (40)

$$k_{0\cdot05} = \frac{7\cdot447}{\sqrt{10}} = 2\cdot355; \quad k_{0\cdot95} = \frac{2\cdot250}{\sqrt{10}} = 0\cdot712.$$

The above calculations have been performed using Table IV and the inverse method described in the previous section. Actually this example could have been treated much more simply by Table V. We shall proceed to show this, but it must be remembered that Table V only covers cases where $\epsilon = 0\cdot05$ and $\epsilon = 0\cdot95$. For other probability levels Table IV will have to be used.

To obtain $t(9, 4\cdot053, 0\cdot05)$ and $t(9, -4\cdot053, 0\cdot05)$ Table V has to be entered with $\eta' = \dfrac{\delta}{\sqrt{(2f)}}\left(1 + \dfrac{\delta^2}{2f}\right)^{-\frac{1}{2}}$, where δ has the two values $\pm 4\cdot053$, i.e. has to be entered with $\eta' = \pm 0\cdot6908$. These give values of λ equal to $1\cdot6800$ and $1\cdot5925$ and are the same as the final λ's obtained in the iterations. We have then only to perform the calculations for t by substituting in (37).

9. Summary

An account of some of the applications of the non-central t-distribution* has been given. Tables of the distribution have been provided together with numerical examples of their use.

If t_0 and δ are connected by the equation $P(t>t_0 \mid \delta) = \epsilon$, then it was found that a good approximation to δ, given t_0, is

$$\delta = t_0 - K_\epsilon\left(1+\frac{t_0^2}{2f}\right)^{\frac{1}{2}},$$

where K_ϵ is the normal deviate exceeded with probability ϵ. The equivalent approximation

$$t_0 = \frac{\delta + K_\epsilon\left(1+\frac{\delta^2}{2f}-\frac{K_\epsilon^2}{2f}\right)^{\frac{1}{2}}}{\left(1-\frac{K_\epsilon^2}{2f}\right)}$$

is good for calculating t_0 given δ. These approximations were seen to have advantages over other approximations also based only on the normal distribution.

Tables IV and V at the end of the paper make possible a more exact determination of δ given t_0 and t_0 given δ, respectively. In these tables a quantity λ_ϵ is given such that

$$\delta = t_0 - \lambda_\epsilon\left(1+\frac{t_0^2}{2f}\right)^{\frac{1}{2}},$$

and so

$$t_0 = \frac{\delta + \lambda_\epsilon\left(1+\frac{\delta^2}{2f}-\frac{\lambda_\epsilon^2}{2f}\right)^{\frac{1}{2}}}{\left(1-\frac{\lambda_\epsilon^2}{2f}\right)}.$$

Note on the accuracy of the tables. In the greater part of the tables λ_ϵ is correct to as many figures as are shown. Occasionally, however, the values of λ_ϵ may be almost 2 units wrong in the last figure. It was nevertheless considered worth while to give all the figures shown.

Throughout Table IV linear interpolation will always give a result not more than a $\frac{1}{2}$ unit wrong in the second last figure. This is also true of Table V, except between $\eta' = -0{\cdot}8$ and $-1{\cdot}0$ for $f = 9$ and 16. Here linear interpolation with respect to η' may be 1 unit wrong in the second last figure.

* $t=(z+\delta)/\sqrt{w}$, where z is a unit normal deviate, w is distributed independently as χ^2/f, and f is the number of degrees of freedom of χ^2.

TABLE IV

Values of $\lambda(f, t_0, \epsilon)$

$\epsilon = 0{\cdot}005$

t_0	y'	y	$f=4$	5	6	7	8	9	16	36	144	∞
								$12/\sqrt{f}=4$	3	2	1	0
$-\infty$	−1·0	0·0	2·62	2·62	2·63	2·63	2·63	2·626	2·622	2·613	2·598	2·576
		0·1	2·62	2·62	2·62	2·62	2·62	2·624	2·620	2·612	2·597	2·576
		0·2	2·61	2·62	2·62	2·62	2·62	2·619	2·616	2·608	2·595	2·576
		0·3	2·60	2·61	2·61	2·61	2·61	2·610	2·609	2·603	2·592	2·576
		0·4	2·59	2·59	2·60	2·60	2·60	2·599	2·599	2·596	2·588	2·576
		0·5	2·57	2·58	2·58	2·58	2·58	2·585	2·588	2·588	2·584	2·576
		0·6	2·55	2·56	2·56	2·57	2·57	2·569	2·575	2·578	2·579	2·576
		0·7	2·53	2·54	2·54	2·55	2·55	2·554	2·562	2·568	2·574	2·576
Negative	−0·6	0·8	2·52	2·52	2·53	2·53	2·54	2·541	2·552	2·561	2·569	2·576
	−0·6	0·8	2·52	2·52	2·53	2·53	2·54	2·541	2·552	2·561	2·569	2·576
	−0·5		2·51	2·52	2·53	2·53	2·53	2·537	2·548	2·558	2·568	2·576
	−0·4		2·52	2·52	2·53	2·53	2·54	2·538	2·548	2·558	2·567	2·576
	−0·3		2·53	2·53	2·54	2·54	2·54	2·544	2·552	2·560	2·568	2·576
	−0·2		2·54	2·54	2·55	2·55	2·55	2·553	2·559	2·565	2·570	2·576
	−0·1		2·56	2·56	2·56	2·56	2·56	2·564	2·567	2·570	2·573	2·576
0	0·0	1·0	2·58	2·58	2·58	2·58	2·58	2·576	2·576	2·576	2·576	2·576
	0·1		2·59	2·59	2·59	2·59	2·59	2·587	2·584	2·581	2·579	2·576
	0·2		2·60	2·60	2·60	2·60	2·60	2·596	2·591	2·586	2·581	2·576
	0·3		2·61	2·61	2·61	2·60	2·60	2·601	2·596	2·590	2·583	2·576
	0·4		2·61	2·61	2·61	2·60	2·60	2·602	2·597	2·591	2·584	2·576
	0·5		2·60	2·60	2·60	2·60	2·60	2·596	2·593	2·589	2·583	2·576
	0·6	0·8	2·57	2·57	2·58	2·58	2·58	2·582	2·584	2·583	2·581	2·576
Positive	0·6	0·8	2·57	2·57	2·58	2·58	2·58	2·582	2·584	2·583	2·581	2·576
		0·7	2·52	2·53	2·54	2·54	2·55	2·552	2·564	2·571	2·575	2·576
		0·6	2·46	2·48	2·49	2·50	2·51	2·518	2·541	2·558	2·569	2·576
		0·5	2·39	2·42	2·45	2·46	2·47	2·484	2·517	2·544	2·563	2·576
		0·4	2·33	2·37	2·40	2·42	2·44	2·451	2·495	2·531	2·557	2·576
		0·3	2·27	2·33	2·36	2·39	2·41	2·422	2·477	2·520	2·552	2·576
		0·2	2·23	2·29	2·33	2·36	2·38	2·399	2·462	2·511	2·548	2·576
		0·1	2·20	2·27	2·31	2·34	2·37	2·385	2·453	2·506	2·546	2·576
∞	1·0	0·0	2·19	2·26	2·30	2·34	2·36	2·380	2·450	2·504	2·545	2·576

$$y = \left(1 + \frac{t_0^2}{2f}\right)^{-\frac{1}{2}}; \qquad y' = \frac{t_0}{\sqrt{(2f)}}\left(1 + \frac{t_0^2}{2f}\right)^{-\frac{1}{2}}.$$

For note on accuracy of table see end of Section 9.

$$\delta(f, t_0, \epsilon) = t_0 - \lambda(f, t_0, \epsilon)\left(1 + \frac{t_0^2}{2f}\right)^{\frac{1}{2}}; \qquad t(f, \delta, \epsilon) = \frac{\delta + \lambda\left(1 + \frac{\delta^2}{2f} - \frac{\lambda^2}{2f}\right)^{\frac{1}{2}}}{\left(1 - \frac{\lambda^2}{2f}\right)}.$$

TABLE IV (*continued*)

Values of $\lambda(f, t_0, \epsilon)$

$\epsilon = 0{\cdot}01$

t_0	y'	y	$f=4$	5	6	7	8	9	16	36	144	∞
								$12/\sqrt{f}=4$	3	2	1	0
$-\infty$	$-1{\cdot}0$	0·0	2·32	2·33	2·33	2·34	2·34	2·340	2·343	2·342	2·337	2·326
		0·1	2·32	2·33	2·33	2·34	2·34	2·339	2·342	2·342	2·337	2·326
		0·2	2·32	2·32	2·33	2·33	2·33	2·335	2·339	2·339	2·335	2·326
		0·3	2·31	2·32	2·32	2·32	2·33	2·328	2·334	2·335	2·333	2·326
		0·4	2·30	2·31	2·31	2·32	2·32	2·320	2·327	2·330	2·330	2·326
		0·5	2·29	2·30	2·30	2·30	2·31	2·310	2·318	2·324	2·327	2·326
Negative		0·6	2·27	2·28	2·29	2·29	2·30	2·299	2·310	2·318	2·323	2·326
		0·7	2·26	2·27	2·28	2·28	2·29	2·289	2·301	2·311	2·320	2·326
	$-0{\cdot}6$	0·8	2·25	2·26	2·27	2·28	2·28	2·282	2·295	2·307	2·317	2·326
	$-0{\cdot}6$	0·8	2·25	2·26	2·27	2·28	2·28	2·282	2·295	2·307	2·317	2·326
	$-0{\cdot}5$		2·26	2·26	2·27	2·28	2·28	2·282	2·294	2·306	2·317	2·326
	$-0{\cdot}4$		2·26	2·27	2·28	2·28	2·28	2·286	2·297	2·307	2·317	2·326
	$-0{\cdot}3$		2·28	2·28	2·29	2·29	2·29	2·293	2·302	2·310	2·319	2·326
	$-0{\cdot}2$		2·29	2·30	2·30	2·30	2·30	2·303	2·309	2·315	2·321	2·326
	$-0{\cdot}1$		2·31	2·31	2·31	2·31	2·31	2·315	2·318	2·320	2·323	2·326
0	0·0	1·0	2·33	2·33	2·33	2·33	2·33	2·326	2·326	2·326	2·326	2·326
	0·1		2·34	2·34	2·34	2·34	2·34	2·337	2·335	2·332	2·329	2·326
	0·2		2·35	2·35	2·35	2·35	2·35	2·347	2·342	2·337	2·332	2·326
	0·3		2·36	2·36	2·36	2·36	2·35	2·353	2·347	2·341	2·334	2·326
	0·4		2·36	2·36	2·36	2·36	2·36	2·356	2·350	2·343	2·335	2·326
	0·5		2·36	2·36	2·36	2·36	2·35	2·353	2·349	2·343	2·335	2·326
	0·6	0·8	2·34	2·34	2·34	2·34	2·34	2·344	2·343	2·340	2·334	2·326
	0·6	0·8	2·34	2·34	2·34	2·34	2·34	2·344	2·343	2·340	2·334	2·326
Positive		0·7	2·30	2·31	2·32	2·32	2·32	2·324	2·329	2·332	2·330	2·326
		0·6	2·26	2·27	2·28	2·29	2·30	2·299	2·313	2·322	2·326	2·326
		0·5	2·21	2·23	2·25	2·26	2·27	2·274	2·296	2·312	2·322	2·326
		0·4	2·16	2·19	2·21	2·23	2·24	2·250	2·280	2·302	2·317	2·326
		0·3	2·12	2·16	2·19	2·20	2·22	2·229	2·267	2·294	2·314	2·326
		0·2	2·09	2·13	2·16	2·19	2·20	2·213	2·256	2·288	2·311	2·326
		0·1	2·07	2·12	2·15	2·17	2·19	2·203	2·250	2·285	2·309	2·326
∞	1·0	0·0	2·06	2·11	2·15	2·17	2·19	2·199	2·247	2·283	2·309	2·326

$$y = \left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}};$$

$$y' = \frac{t_0}{\sqrt{(2f)}}\left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}}.$$

For note on accuracy of table see end of Section 9.

$$\delta(f, t_0, \epsilon) = t_0 - \lambda(f, t_0, \epsilon)\left(1+\frac{t_0^2}{2f}\right)^{\frac{1}{2}}; \qquad t(f, \delta, \epsilon) = \frac{\delta + \lambda\left(1+\frac{\delta^2}{2f}-\frac{\lambda^2}{2f}\right)^{\frac{1}{2}}}{\left(1-\frac{\lambda^2}{2f}\right)}.$$

TABLE IV (*continued*)

Values of $\lambda(f, t_0, \epsilon)$

$\epsilon = 0{\cdot}025$

t_0	y'	y	$f=4$	5	6	7	8	9	16	36	144	∞
								$12/\sqrt{f}=4$	3	2	1	0
−∞	−1·0	0·0	1·892	1·904	1·912	1·917	1·922	1·925	1·939	1·950	1·956	1·960
		0·1	1·891	1·903	1·911	1·917	1·921	1·925	1·938	1·949	1·956	1·960
		0·2	1·889	1·900	1·908	1·914	1·919	1·923	1·936	1·947	1·955	1·960
		0·3	1·885	1·897	1·905	1·911	1·916	1·919	1·934	1·945	1·954	1·960
		0·4	1·880	1·892	1·900	1·907	1·912	1·915	1·930	1·943	1·953	1·960
		0·5	1·875	1·887	1·896	1·902	1·907	1·911	1·926	1·940	1·951	1·960
		0·6	1·871	1·883	1·891	1·898	1·903	1·907	1·923	1·937	1·949	1·960
		0·7	1·869	1·880	1·889	1·895	1·900	1·904	1·920	1·935	1·948	1·960
Negative	−0·6	0·8	1·872	1·883	1·890	1·896	1·901	1·905	1·920	1·934	1·948	1·960
	−0·6	0·8	1·872	1·883	1·890	1·896	1·901	1·905	1·920	1·934	1·948	1·960
	−0·5		1·881	1·890	1·896	1·902	1·906	1·909	1·923	1·936	1·948	1·960
	−0·4		1·893	1·900	1·906	1·910	1·913	1·916	1·928	1·939	1·950	1·960
	−0·3		1·908	1·913	1·918	1·921	1·924	1·926	1·935	1·943	1·952	1·960
	−0·2		1·925	1·928	1·931	1·933	1·935	1·937	1·943	1·948	1·954	1·960
	−0·1		1·942	1·944	1·946	1·947	1·948	1·948	1·951	1·954	1·957	1·960
0	0·0	1·0	1·960	1·960	1·960	1·960	1·960	1·960	1·960	1·960	1·960	1·960
	0·1		1·976	1·975	1·974	1·973	1·972	1·971	1·968	1·966	1·963	1·960
	0·2		1·990	1·987	1·985	1·984	1·982	1·981	1·976	1·971	1·966	1·960
	0·3		2·000	1·997	1·994	1·992	1·991	1·989	1·982	1·975	1·968	1·960
	0·4		2·006	2·003	2·000	1·998	1·996	1·994	1·987	1·979	1·970	1·960
	0·5		2·006	2·004	2·001	2·000	1·998	1·996	1·989	1·981	1·971	1·960
	0·6	0·8	1·998	1·998	1·997	1·996	1·995	1·994	1·988	1·980	1·971	1·960
Positive	0·6	0·8	1·998	1·998	1·997	1·996	1·995	1·994	1·988	1·980	1·971	1·960
		0·7	1·978	1·982	1·983	1·984	1·984	1·985	1·983	1·978	1·970	1·960
		0·6	1·953	1·961	1·966	1·969	1·971	1·972	1·975	1·973	1·968	1·960
		0·5	1·926	1·939	1·947	1·952	1·956	1·959	1·966	1·968	1·966	1·960
		0·4	1·900	1·918	1·929	1·936	1·942	1·946	1·957	1·963	1·964	1·960
		0·3	1·878	1·900	1·913	1·923	1·930	1·935	1·950	1·959	1·962	1·960
		0·2	1·860	1·885	1·901	1·912	1·920	1·926	1·945	1·956	1·960	1·960
		0·1	1·849	1·876	1·893	1·906	1·914	1·921	1·941	1·954	1·959	1·960
∞	1·0	0·0	1·844	1·872	1·890	1·903	1·912	1·919	1·940	1·953	1·959	1·960

$$y = \left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}};$$

$$y' = \frac{t_0}{\sqrt{(2f)}}\left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}}.$$

For note on accuracy of table see end of Section 9.

$$\delta(f, t_0, \epsilon) = t_0 - \lambda(f, t_0, \epsilon)\left(1+\frac{t_0^2}{2f}\right)^{\frac{1}{2}}; \qquad t(f, \delta, \epsilon) = \frac{\delta + \lambda\left(1+\frac{\delta^2}{2f}-\frac{\lambda^2}{2f}\right)^{\frac{1}{2}}}{\left(1-\frac{\lambda^2}{2f}\right)}.$$

TABLE IV (*continued*)

Values of $\lambda(f, t_0, \epsilon)$

$\epsilon = 0{\cdot}05$

t_0	y'	y	$f=4$	5	6	7	8	9	16	36	144	∞
								$12/\sqrt{f}=4$	3	2	1	0
$-\infty$	$-1{\cdot}0$	0·0	1·528	1·543	1·554	1·563	1·569	1·5744	1·5952	1·6141	1·6307	1·6449
		0·1	1·527	1·543	1·554	1·562	1·569	1·5742	1·5950	1·6139	1·6306	1·6449
		0·2	1·527	1·542	1·553	1·562	1·568	1·5736	1·5945	1·6135	1·6303	1·6449
		0·3	1·526	1·542	1·553	1·561	1·567	1·5728	1·5937	1·6129	1·6300	1·6449
		0·4	1·526	1·541	1·552	1·560	1·567	1·5720	1·5930	1·6122	1·6295	1·6449
		0·5	1·526	1·541	1·552	1·560	1·566	1·5717	1·5924	1·6116	1·6291	1·6449
		0·6	1·529	1·543	1·553	1·561	1·567	1·5724	1·5926	1·6115	1·6289	1·6449
Negative		0·7	1·534	1·548	1·557	1·565	1·570	1·5751	1·5942	1·6123	1·6292	1·6449
	$-0{\cdot}6$	0·8	1·546	1·557	1·566	1·572	1·577	1·5816	1·5986	1·6149	1·6303	1·6449
	$-0{\cdot}6$	0·8	1·546	1·557	1·566	1·572	1·577	1·5816	1·5986	1·6149	1·6303	1·6449
	$-0{\cdot}5$		1·559	1·569	1·576	1·581	1·586	1·5895	1·6041	1·6183	1·6319	1·6449
	$-0{\cdot}4$		1·575	1·582	1·588	1·593	1·596	1·5990	1·6110	1·6226	1·6339	1·6449
	$-0{\cdot}3$		1·592	1·597	1·602	1·605	1·608	1·6097	1·6187	1·6276	1·6363	1·6449
	$-0{\cdot}2$		1·609	1·613	1·616	1·618	1·620	1·6212	1·6272	1·6331	1·6390	1·6449
	$-0{\cdot}1$		1·627	1·629	1·630	1·632	1·632	1·6332	1·6360	1·6390	1·6419	1·6449
0	0·0	1·0	1·645	1·645	1·645	1·645	1·645	1·6449	1·6449	1·6449	1·6449	1·6449
	0·1		1·661	1·660	1·658	1·658	1·657	1·6561	1·6534	1·6506	1·6478	1·6449
	0·2		1·676	1·673	1·671	1·669	1·668	1·6665	1·6614	1·6561	1·6506	1·6449
	0·3		1·688	1·684	1·681	1·679	1·677	1·6756	1·6686	1·6610	1·6531	1·6449
	0·4		1·697	1·693	1·690	1·687	1·685	1·6830	1·6745	1·6652	1·6553	1·6449
	0·5		1·702	1·698	1·695	1·692	1·690	1·6881	1·6789	1·6685	1·6571	1·6449
	0·6	0·8	1·703	1·700	1·697	1·695	1·693	1·6906	1·6815	1·6707	1·6584	1·6449
Positive	0·6	0·8	1·703	1·700	1·697	1·695	1·693	1·6906	1·6815	1·6707	1·6584	1·6449
		0·7	1·696	1·695	1·694	1·692	1·691	1·6895	1·6817	1·6714	1·6590	1·6449
		0·6	1·685	1·687	1·687	1·687	1·686	1·6855	1·6797	1·6706	1·6588	1·6449
		0·5	1·673	1·677	1·680	1·680	1·680	1·6803	1·6767	1·6691	1·6583	1·6449
		0·4	1·661	1·668	1·672	1·674	1·675	1·6750	1·6735	1·6674	1·6576	1·6449
		0·3	1·651	1·660	1·665	1·668	1·670	1·6704	1·6706	1·6658	1·6569	1·6449
		0·2	1·643	1·654	1·660	1·663	1·666	1·6669	1·6684	1·6645	1·6564	1·6449
		0·1	1·639	1·650	1·657	1·661	1·663	1·6646	1·6670	1·6637	1·6560	1·6449
∞	1·0	0·0	1·636	1·648	1·655	1·660	1·662	1·6638	1·6665	1·6634	1·6559	1·6449

$$y = \left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}};$$

$$y' = \frac{t_0}{\sqrt{(2f)}}\left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}}.$$

For note on accuracy of table see end of Section 9.

$$\delta(f, t_0, \epsilon) = t_0 - \lambda(f, t_0, \epsilon)\left(1+\frac{t_0^2}{2f}\right)^{\frac{1}{2}}; \qquad t(f, \delta, \epsilon) = \frac{\delta + \lambda\left(1+\frac{\delta^2}{2f}-\frac{\lambda^2}{2f}\right)^{\frac{1}{2}}}{\left(1-\frac{\lambda^2}{2f}\right)}.$$

TABLE IV (*continued*)

Values of $\lambda(f, t_0, \epsilon)$

$\epsilon = 0{\cdot}10$

t_0	y'	y	$f=4$	5	6	7	8	9	16	36	144	∞
								$12/\sqrt{f}=4$	3	2	1	0
$-\infty$	−1·0	0·0	1·116	1·136	1·150	1·161	1·169	1·1765	1·2049	1·2319	1·2575	1·2816
		0·1	1·116	1·136	1·150	1·161	1·170	1·1768	1·2051	1·2321	1·2576	1·2816
		0·2	1·118	1·137	1·151	1·162	1·171	1·1777	1·2057	1·2325	1·2578	1·2816
		0·3	1·121	1·140	1·154	1·164	1·172	1·1793	1·2069	1·2332	1·2581	1·2816
		0·4	1·125	1·143	1·157	1·167	1·175	1·1819	1·2087	1·2343	1·2586	1·2816
		0·5	1·131	1·149	1·162	1·171	1·179	1·1856	1·2114	1·2360	1·2594	1·2816
		0·6	1·140	1·157	1·169	1·178	1·185	1·1912	1·2153	1·2385	1·2606	1·2816
Negative		0·7	1·153	1·168	1·179	1·187	1·194	1·1992	1·2210	1·2421	1·2623	1·2816
	−0·6	0·8	1·173	1·185	1·194	1·201	1·206	1·2110	1·2295	1·2475	1·2649	1·2816
	−0·6	0·8	1·173	1·185	1·194	1·201	1·206	1·2110	1·2295	1·2475	1·2649	1·2816
	−0·5		1·191	1·201	1·208	1·214	1·218	1·2222	1·2376	1·2527	1·2673	1·2816
	−0·4		1·209	1·217	1·223	1·227	1·231	1·2338	1·2461	1·2582	1·2700	1·2816
	−0·3		1·228	1·233	1·238	1·241	1·244	1·2458	1·2548	1·2639	1·2728	1·2816
	−0·2		1·246	1·250	1·253	1·255	1·256	1·2578	1·2638	1·2697	1·2757	1·2816
	−0·1		1·264	1·266	1·267	1·268	1·269	1·2698	1·2727	1·2756	1·2786	1·2816
0	0·0	1·0	1·282	1·282	1·282	1·282	1·282	1·2816	1·2816	1·2816	1·2816	1·2816
	0·1		1·298	1·297	1·295	1·294	1·294	1·2929	1·2902	1·2874	1·2845	1·2816
	0·2		1·313	1·310	1·308	1·306	1·305	1·3038	1·2985	1·2930	1·2873	1·2816
	0·3		1·327	1·323	1·320	1·318	1·316	1·3140	1·3064	1·2984	1·2901	1·2816
	0·4		1·340	1·335	1·331	1·328	1·326	1·3233	1·3137	1·3035	1·2927	1·2816
	0·5		1·350	1·345	1·341	1·337	1·334	1·3316	1·3204	1·3082	1·2952	1·2816
	0·6	0·8	1·358	1·353	1·349	1·345	1·342	1·3387	1·3263	1·3124	1·2974	1·2816
Positive	0·6	0·8	1·358	1·353	1·349	1·345	1·342	1·3387	1·3263	1·3124	1·2974	1·2816
		0·7	1·365	1·360	1·355	1·352	1·348	1·3453	1·3320	1·3166	1·2997	1·2816
		0·6	1·367	1·363	1·359	1·355	1·352	1·3492	1·3354	1·3193	1·3012	1·2816
		0·5	1·368	1·365	1·361	1·358	1·354	1·3516	1·3377	1·3210	1·3021	1·2816
		0·4	1·369	1·366	1·362	1·359	1·356	1·3531	1·3392	1·3222	1·3028	1·2816
		0·3	1·369	1·367	1·363	1·360	1·357	1·3541	1·3401	1·3229	1·3032	1·2816
		0·2	1·370	1·367	1·364	1·361	1·358	1·3548	1·3408	1·3234	1·3035	1·2816
		0·1	1·370	1·367	1·364	1·361	1·358	1·3552	1·3411	1·3237	1·3037	1·2816
∞	1·0	0·0	1·370	1·368	1·364	1·361	1·358	1·3554	1·3413	1·3238	1·3038	1·2816

$$y = \left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}};$$

$$y' = \frac{t_0}{\sqrt{(2f)}}\left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}}.$$

For note on accuracy of table see end of Section 9.

$$\delta(f, t_0, \epsilon) = t_0 - \lambda(f, t_0, \epsilon)\left(1+\frac{t_0^2}{2f}\right)^{\frac{1}{2}}; \qquad t(f, \delta, \epsilon) = \frac{\delta + \lambda\left(1+\frac{\delta^2}{2f}-\frac{\lambda^2}{2f}\right)^{\frac{1}{2}}}{\left(1-\frac{\lambda^2}{2f}\right)}.$$

TABLE IV (*continued*)

Values of $\lambda(f, t_0, \epsilon)$

$\epsilon = 0{\cdot}20$

t_0	y'	y	$f=4$	5	6	7	8	9	16	36	144	∞
								$12/\sqrt{f}=4$	3	2	1	0
$-\infty$	$-1{\cdot}0$	0·0	·632	·656	·673	·686	·697	·7055	·7408	·7753	·8089	·8416
		0·1	·633	·657	·674	·687	·698	·7063	·7414	·7756	·8091	·8416
		0·2	·637	·660	·677	·690	·700	·7086	·7431	·7768	·8096	·8416
		0·3	·643	·665	·682	·694	·704	·7125	·7460	·7787	·8106	·8416
		0·4	·652	·673	·689	·701	·710	·7182	·7503	·7815	·8120	·8416
		0·5	·664	·684	·698	·710	·718	·7259	·7559	·7853	·8138	·8416
		0·6	·679	·697	·711	·721	·729	·7359	·7633	·7901	·8162	·8416
		0·7	·699	·715	·727	·736	·743	·7487	·7727	·7962	·8192	·8416
Negative	$-0{\cdot}6$	0·8	·725	·738	·747	·755	·760	·7652	·7848	·8041	·8231	·8416
	$-0{\cdot}6$	0·8	·725	·738	·747	·755	·760	·7652	·7848	·8041	·8231	·8416
	$-0{\cdot}5$		·747	·757	·765	·771	·775	·7791	·7950	·8108	·8264	·8416
	$-0{\cdot}4$		·768	·776	·781	·786	·790	·7925	·8049	·8173	·8295	·8416
	$-0{\cdot}3$		·788	·793	·797	·801	·803	·8054	·8145	·8236	·8326	·8416
	$-0{\cdot}2$		·806	·810	·813	·815	·816	·8178	·8238	·8297	·8357	·8416
	$-0{\cdot}1$		·824	·826	·827	·828	·829	·8299	·8328	·8357	·8387	·8416
0	0·0	1·0	·842	·842	·842	·842	·842	·8416	·8416	·8416	·8416	·8416
	0·1		·858	·857	·856	·855	·854	·8531	·8503	·8475	·8446	·8416
	0·2		·875	·872	·869	·867	·866	·8644	·8589	·8533	·8475	·8416
	0·3		·890	·886	·883	·880	·878	·8756	·8675	·8591	·8505	·8416
	0·4		·906	·900	·896	·892	·889	·8868	·8762	·8650	·8535	·8416
	0·5		·921	·914	·909	·905	·901	·8980	·8849	·8710	·8565	·8416
	0·6	0·8	·937	·929	·922	·917	·913	·9094	·8937	·8771	·8596	·8416
Positive	0·6	0·8	·937	·929	·922	·917	·913	·9094	·8937	·8771	·8596	·8416
		0·7	·955	·946	·938	·932	·927	·9229	·9042	·8843	·8633	·8416
		0·6	·970	·959	·951	·944	·939	·9337	·9125	·8899	·8662	·8416
		0·5	·982	·971	·961	·954	·948	·9424	·9192	·8945	·8685	·8416
		0·4	·992	·980	·970	·962	·955	·9495	·9246	·8981	·8703	·8416
		0·3	1·001	·987	·977	·968	·961	·9550	·9288	·9008	·8717	·8416
		0·2	1·007	·993	·982	·973	·965	·9591	·9317	·9028	·8726	·8416
		0·1	1·011	·996	·985	·976	·968	·9615	·9335	·9040	·8732	·8416
∞	1·0	0·0	1·012	·997	·986	·977	·969	·9623	·9341	·9044	·8734	·8416

$$y = \left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}};$$

$$y' = \frac{t_0}{\sqrt{(2f)}}\left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}}.$$

For note on accuracy of table see end of Section 9.

$$\delta(f, t_0, \epsilon) = t_0 - \lambda(f, t_0, \epsilon)\left(1+\frac{t_0^2}{2f}\right)^{\frac{1}{2}}; \qquad t(f, \delta, \epsilon) = \frac{\delta+\lambda\left(1+\frac{\delta^2}{2f}-\frac{\lambda^2}{2f}\right)^{\frac{1}{2}}}{\left(1-\frac{\lambda^2}{2f}\right)}.$$

TABLE IV (*continued*)

Values of $\lambda(f, t_0, \epsilon)$

$\epsilon = 0{\cdot}30$

t_0	y'	y	$f=4$	5	6	7	8	9	16	36	144	∞
								$12/\sqrt{f}=4$	3	2	1	0
$-\infty$	$-1{\cdot}0$	0·0	·295	·320	·339	·353	·365	·3739	·4124	·4503	·4876	·5244
		0·1	·297	·322	·340	·354	·366	·3749	·4132	·4508	·4878	·5244
		0·2	·301	·326	·344	·358	·369	·3780	·4154	·4523	·4886	·5244
		0·3	·309	·333	·350	·364	·374	·3831	·4193	·4548	·4899	·5244
		0·4	·320	·343	·359	·372	·382	·3904	·4247	·4584	·4917	·5244
		0·5	·335	·356	·371	·383	·392	·4000	·4318	·4632	·4940	·5244
Negative		0·6	·354	·373	·386	·397	·405	·4122	·4408	·4691	·4970	·5244
		0·7	·377	·393	·405	·414	·421	·4272	·4519	·4764	·5006	·5244
	$-0{\cdot}6$	0·8	·406	·419	·428	·435	·441	·4458	·4656	·4854	·5050	·5244
	$-0{\cdot}6$	0·8	·406	·419	·428	·435	·441	·4458	·4656	·4854	·5050	·5244
	$-0{\cdot}5$		·429	·439	·446	·452	·457	·4608	·4768	·4928	·5087	·5244
	$-0{\cdot}4$		·451	·458	·464	·468	·472	·4749	·4873	·4997	·5121	·5244
	$-0{\cdot}3$		·471	·476	·480	·483	·486	·4881	·4971	·5062	·5153	·5244
	$-0{\cdot}2$		·489	·493	·496	·498	·499	·5007	·5065	·5125	·5184	·5244
	$-0{\cdot}1$		·507	·509	·510	·511	·512	·5127	·5156	·5185	·5214	·5244
0	0·0	1·0	·524	·524	·524	·524	·524	·5244	·5244	·5244	·5244	·5244
	0·1		·541	·540	·538	·537	·537	·5360	·5331	·5303	·5273	·5244
	0·2		·558	·555	·552	·550	·549	·5475	·5419	·5362	·5303	·5244
	0·3		·575	·570	·567	·564	·561	·5593	·5509	·5422	·5334	·5244
	0·4		·592	·586	·581	·577	·574	·5715	·5602	·5485	·5366	·5244
	0·5		·610	·602	·596	·591	·587	·5842	·5699	·5551	·5399	·5244
	0·6	0·8	·630	·620	·613	·607	·602	·5977	·5803	·5621	·5434	·5244
Positive	0·6	0·8	·630	·620	·613	·607	·602	·5977	·5803	·5621	·5434	·5244
		0·7	·654	·642	·633	·626	·620	·6146	·5931	·5707	·5478	·5244
		0·6	·674	·660	·650	·641	·634	·6285	·6036	·5778	·5513	·5244
		0·5	·691	·676	·664	·654	·647	·6401	·6123	·5836	·5542	·5244
		0·4	·706	·689	·676	·665	·657	·6496	·6194	·5883	·5566	·5244
		0·3	·717	·699	·685	·674	·665	·6570	·6249	·5920	·5584	·5244
		0·2	·726	·706	·692	·680	·670	·6624	·6289	·5946	·5597	·5244
		0·1	·731	·711	·696	·684	·674	·6657	·6312	·5961	·5604	·5244
∞	1·0	0·0	·733	·712	·697	·685	·675	·6668	·6320	·5966	·5607	·5244

$$y = \left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}};$$

$$y' = \frac{t_0}{\sqrt{(2f)}}\left(1+\frac{t_0^2}{2f}\right)^{-\frac{1}{2}}.$$

For note on accuracy of table see end of Section 9.

$$\delta(f, t_0, \epsilon) = t_0 - \lambda(f, t_0, \epsilon)\left(1+\frac{t_0^2}{2f}\right)^{\frac{1}{2}}; \qquad t(f, \delta, \epsilon) = \frac{\delta+\lambda\left(1+\frac{\delta^2}{2f}-\frac{\lambda^2}{2f}\right)^{\frac{1}{2}}}{\left(1-\frac{\lambda^2}{2f}\right)}.$$

TABLE IV (*continued*)

Values of $\lambda(f, t_0, \epsilon)$

$\epsilon = 0{\cdot}40$

t_0	y'	y	$f=4$	5	6	7	8	9	16	36	144	∞
								$12/\sqrt{f}=4$	3	2	1	0
$-\infty$	$-1{\cdot}0$	0·0	·016	·041	·060	·075	·087	·0964	·1362	·1755	·2146	·2533
		0·1	·018	·043	·062	·076	·088	·0975	·1370	·1761	·2149	·2533
		0·2	·023	·048	·066	·080	·092	·1010	·1396	·1778	·2157	·2533
		0·3	·032	·056	·073	·087	·098	·1068	·1439	·1807	·2172	·2533
		0·4	·044	·067	·083	·096	·106	·1150	·1500	·1847	·2192	·2533
		0·5	·061	·081	·096	·108	·118	·1256	·1578	·1899	·2218	·2533
		0·6	·081	·099	·113	·123	·132	·1388	·1676	·1964	·2250	·2533
Negative		0·7	·106	·121	·133	·142	·149	·1548	·1795	·2042	·2289	·2533
	$-0{\cdot}6$	0·8	·136	·148	·157	·164	·169	·1742	·1940	·2138	·2336	·2533
	$-0{\cdot}6$	0·8	·136	·148	·157	·164	·169	·1742	·1940	·2138	·2336	[illegible]
	$-0{\cdot}5$		·159	·169	·176	·181	·186	·1897	·2055	·2215	·2374	[illegible]
	$-0{\cdot}4$		·180	·188	·193	·197	·201	·2039	·2162	·2285	·2409	[illegible]
	$-0{\cdot}3$		·200	·205	·209	·212	·215	·2172	·2261	·2352	·2442	[illegible]
	$-0{\cdot}2$		·219	·222	·225	·227	·228	·2297	·2355	·2414	·2474	[illegible]
	$-0{\cdot}1$		·236	·238	·239	·240	·241	·2417	·2445	·2474	·2504	[illegible]
0	0·0	1·0	·253	·253	·253	·253	·253	·2533	·2533	·2533	·2533	[illegible]
	0·1		·270	·269	·267	·266	·266	·2649	·2621	·2592	·2563	[illegible]
	0·2		·287	·284	·282	·280	·278	·2767	·2710	·2652	·2593	[illegible]
	0·3		·305	·300	·296	·293	·291	·2888	·2802	·2714	·2624	[illegible]
	0·4		·323	·317	·312	·308	·304	·3016	·2899	·2779	·2657	[illegible]
	0·5		·343	·335	·328	·323	·319	·3152	·3002	·2848	·2692	[illegible]
	0·6	0·8	·365	·354	·346	·340	·334	·3299	·3113	·2923	·2729	·2533
Positive	0·6	0·8	·365	·354	·346	·340	·334	·3299	·3113	·2923	·2729	·2533
		0·7	·392	·379	·369	·361	·354	·3485	·3254	·3017	·2776	·2533
		0·6	·416	·400	·388	·378	·371	·3641	·3370	·3094	·2815	·2533
		0·5	·435	·417	·404	·393	·384	·3771	·3467	·3158	·2847	·2533
		0·4	·452	·432	·417	·405	·396	·3877	·3546	·3211	·2873	·2533
		0·3	·465	·443	·427	·415	·404	·3960	·3607	·3251	·2893	·2533
		0·2	·474	·452	·435	·422	·411	·4019	·3651	·3280	·2907	·2533
		0·1	·480	·457	·439	·426	·415	·405[illegible]	·3677	·3297	·2916	·2533
∞	1·0	0·0	·482	·458	·441	·427	·416	·4067	·3686	·3303	·2919	·2533

$$y = \left(1 + \frac{t_0^2}{2f}\right)^{-\frac{1}{2}};$$

$$y' = \frac{t_0}{\sqrt{(2f)}}\left(1 + \frac{t_0^2}{2f}\right)^{-\frac{1}{2}}$$

For note on accuracy of table see end of Section 9.

$$\delta(f, t_0, \epsilon) = t_0 - \lambda(f, t_0, \epsilon)\left(1 + \frac{t_0^2}{2f}\right)^{\frac{1}{2}}; \qquad t(f, \delta, \epsilon) = \frac{\delta + \lambda\left(1 + \frac{\delta^2}{2f} - \frac{\lambda^2}{2f}\right)^{\frac{1}{2}}}{\left(1 - \frac{\lambda^2}{2f}\right)}.$$

TABLE IV (*continued*)

Values of $\lambda(f,\, t_0,\, \epsilon)$

$\epsilon = 0{\cdot}50$

t_0	y'	y	$f=4$	5	6	7	8	9	16	36	144	∞
								$12/\sqrt{f}=4$	3	2	1	0
$-\infty$	−1·0	0·0	−·237	−·212	−·194	−·179	−·167	−·1578	−·1182	−·0787	−·0393	·0000
		0·1	−·235	−·210	−·192	−·178	−·166	−·1566	−·1173	−·0781	−·0390	·000C
		0·2	−·230	−·205	−·188	−·174	−·162	−·1530	−·1146	−·0763	−·0381	·0000
		0·3	−·220	−·197	−·180	−·167	−·156	−·1469	−·1101	−·0733	−·0366	·0000
		0·4	−·207	−·185	−·169	−·157	−·147	−·1384	−·1038	−·0692	−·0345	·0000
		0·5	−·190	−·171	−·156	−·144	−·135	−·1275	−·0957	−·0638	−·0319	·0000
		0·6	−·169	−·152	−·139	−·129	−·121	−·1140	−·0857	−·0572	−·0286	·0000
Negative		0·7	−·145	−·130	−·119	−·111	−·104	−·0979	−·0736	−·0492	−·0246	·0000
	−0·6	0·8	−·116	−·104	−·095	−·089	−·083	−·0784	−·0591	−·0395	−·0198	·0000
	−0·6	0·8	−·116	−·104	−·095	−·089	−·083	−·0784	−·0591	−·0395	−·0198	·0000
	−0·5		−·093	−·084	−·077	−·071	−·067	−·0631	−·0476	−·0318	−·0159	·0000
	−0·4		−·072	−·065	−·060	−·055	−·052	−·0490	−·0370	−·0248	−·0124	·0000
	−0·3		−·053	−·047	−·044	−·041	−·038	−·0359	−·0271	−·0181	−·0091	·0000
	−0·2		−·034	−·031	−·029	−·027	−·025	−·0235	−·0178	−·0119	−·0060	·0000
	−0·1		−·017	−·015	−·014	−·013	−·012	−·0116	−·0088	−·0059	−·0030	·0000
0	0·0	1·0	·000	·000	·000	·000	·000	·0000	·0000	·0000	·0000	·0000
	0·1		·017	·015	·014	·013	·012	·0116	·0088	·0059	·0030	·0000
	0·2		·034	·031	·029	·027	·025	·0235	·0178	·0119	·0060	·0000
	0·3		·053	·047	·044	·041	·038	·0359	·0271	·0181	·0091	·0000
	0·4		·072	·065	·060	·055	·052	·0490	·0370	·0248	·0124	·0000
	0·5		·093	·084	·077	·071	·067	·0631	·0476	·0318	·0159	·0000
	0·6	0·8	·116	·104	·095	·089	·083	·0784	·0591	·0395	·0198	·0000
	0·6	0·8	·116	·104	·095	·089	·083	·0784	·0591	·0395	·0198	·0000
Positive		0·7	·145	·130	·119	·111	·104	·0979	·0736	·0492	·0246	·0000
		0·6	·169	·152	·139	·129	·121	·1140	·0857	·0572	·0286	·0000
		0·5	·190	·171	·156	·144	·135	·1275	·0957	·0638	·0319	·0000
		0·4	·207	·185	·169	·157	·147	·1384	·1038	·0692	·0345	·0000
		0·3	·220	·197	·180	·167	·156	·1469	·1101	·0733	·0366	·0000
		0·2	·230	·205	·188	·174	·162	·1530	·1146	·0763	·0381	·0000
		0·1	·235	·210	·192	·178	·166	·1566	·1173	·0781	·0390	·0000
∞	1·0	0·0	·237	·212	·194	·179	·167	·1578	·1182	·0787	·0393	·0000

$$y = \left(1 + \frac{t_0^2}{2f}\right)^{-\frac{1}{2}};$$

$$y' = \frac{t_0}{\sqrt{(2f)}}\left(1 + \frac{t_0^2}{2f}\right)^{-\frac{1}{2}}.$$

For note on accuracy of table see end of Section 9.

$$\delta(f, t_0, \epsilon) = t_0 - \lambda(f, t_0, \epsilon)\left(1 + \frac{t_0^2}{2f}\right)^{\frac{1}{2}}; \qquad t(f, \delta, \epsilon) = \frac{\delta + \lambda\left(1 + \frac{\delta^2}{2f} - \frac{\lambda^2}{2f}\right)^{\frac{1}{2}}}{\left(1 - \frac{\lambda^2}{2f}\right)}.$$

TABLE V

Values of λ as a function of δ

$\epsilon = 0{\cdot}05$

η' \ f	4	5	6	7	8	9	16	36	144	∞
1·0	1·636	1·648	1·655	1·660	1·662	1·6638	1·6665	1·6634	1·6559	1·6449
0·9	1·643	1·655	1·662	1·666	1·668	1·6695	1·6711	1·6667	1·6576	1·6449
0·8	1·650	1·662	1·668	1·672	1·674	1·6747	1·6751	1·6691	1·6586	1·6449
0·7	1·657	1·668	1·674	1·677	1·679	1·6796	1·6782	1·6707	1·6589	1·6449
0·6	1·664	1·675	1·680	1·682	1·684	1·6838	1·6804	1·6714	1·6587	1·6449
0·5	1·671	1·681	1·686	1·687	1·687	1·6871	1·6817	1·6709	1·6580	1·6449
0·4	1·679	1·687	1·690	1·691	1·691	1·6896	1·6816	1·6698	1·6568	1·6449
0·3	1·687	1·693	1·694	1·693	1·692	1·6902	1·6804	1·6677	1·6550	1·6449
0·2	1·693	1·697	1·696	1·694	1·692	1·6898	1·6779	1·6646	1·6529	1·6449
0·1	1·698	1·699	1·697	1·693	1·690	1·6874	1·6738	1·6606	1·6504	1·6449
0·0	1·703	1·699	1·695	1·690	1·686	1·6827	1·6682	1·6558	1·6477	1·6449
−0·1	1·702	1·695	1·689	1·684	1·679	1·6756	1·6611	1·6503	1·6447	1·6449
−0·2	1·698	1·688	1·680	1·674	1·669	1·6657	1·6525	1·6442	1·6417	1·6449
−0·3	1·687	1·676	1·667	1·661	1·657	1·6535	1·6427	1·6378	1·6388	1·6449
−0·4	1·670	1·658	1·650	1·645	1·642	1·6391	1·6322	1·6314	1·6359	1·6449
−0·5	1·646	1·636	1·630	1·627	1·624	1·6231	1·6213	1·6252	1·6334	1·6449
−0·6	1·615	1·610	1·607	1·606	1·606	1·6066	1·6108	1·6195	1·6313	1·6449
−0·7	1·582	1·583	1·585	1·587	1·589	1·5911	1·6019	1·6150	1·6299	1·6449
−0·8	1·551	1·559	1·565	1·571	1·575	1·5792	1·5954	1·6122	1·6291	1·6449
−0·9	1·531	1·544	1·553	1·561	1·567	1·5722	1·5925	1·6116	1·6292	1·6449
−1·0	1·528	1·543	1·554	1·563	1·569	1·5744	1·5952	1·6141	1·6307	1·6449

$$\eta' = \frac{\delta}{\sqrt{(2f)}}\left(1+\frac{\delta^2}{2f}\right)^{-\frac{1}{2}}; \qquad t(f,\delta,\epsilon) = \frac{\delta+\lambda\left(1+\dfrac{\delta^2}{2f}-\dfrac{\lambda^2}{2f}\right)^{\frac{1}{2}}}{\left(1-\dfrac{\lambda^2}{2f}\right)}.$$

For note on accuracy of table see end of Section 9.

APPENDIX

On the calculation of Table IV

Before deciding on the method to be employed in calculating the tables a wide variety of methods were considered. We examined these with a view to finding one which would best combine the attributes of speediness, lack of opportunity for numerical errors, and adaptability to moderately large-scale work. In this appendix will be described the more promising of the methods tested, including the one finally adopted.

The non-central t being defined as in (1) we have for the probability that $t > t_0$, with the notation of Section 2,

$$P(f,\delta,t_0) = \frac{1}{2^{\frac{1}{2}f-1}\,\Gamma(\frac{1}{2}f)}\int_0^\infty \left\{\chi^{f-1}e^{-\frac{1}{2}\chi^2}\frac{1}{\sqrt{(2\pi)}}\int_{\frac{t_0\chi}{\sqrt{f}}}^\infty e^{-\frac{1}{2}(u-\delta)^2}\,du\right\}d\chi. \qquad \text{......(41)}$$

We want to find pairs of values δ and t_0 such that $P(f,\delta,t_0)$ has specified values ϵ; i.e. we want to evaluate the quantities $t_0(f,\delta,\epsilon)$ and $\delta(f,t_0,\epsilon)$.

(i) *Direct Methods*

These are based on the idea of evaluating the right-hand side of (41) directly for various values of δ, f, and t_0 and then obtaining the quantities required by a process of inverse interpolation. An obvious way to evaluate $P(f, \delta, t_0)$ is by quadrature. For small even values of f, however, the following formulae:

$$P(0<t<\infty \mid f, \delta) = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}(u-\delta)^2} du,$$

$$P(0<t<t_0 \mid f, \delta) = \frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}\delta^2(1+f^{-1}t_0^2)^{-1}\} \sum_{r=0}^{\frac{1}{2}(f-2)} \frac{(2r)!}{(1+f^{-1}t_0^2)^r \, 2^r \, r!} Hh_{2r}\{-\delta t_0(f+t_0^2)^{-\frac{1}{2}}\},$$

where $$Hh_s(x) = \frac{1}{s!}\int_0^\infty u^s e^{-\frac{1}{2}(u+x)^2} du$$

were found to be more convenient to use. The Hh_s functions have been tabled by J. R. Airey (1931).

This modified procedure was still unduly lengthy and not well adapted to large scale work. It provides, however, a useful check on values obtained by other methods.

(ii) *Solution by means of a differential equation*

For fixed values of f and ϵ the equation $P(f, \delta, t_0) = \epsilon$ may be regarded as specifying δ and t_0 as implicit functions of each other. This implies that for constant ϵ,

$$\frac{dt_0}{d\delta} = -\left\{\frac{\partial P(f, \delta, t_0)}{\partial \delta} \Big/ \frac{\partial P(f, \delta, t_0)}{\partial t_0}\right\}.$$

From (41) this is found to give

$$\frac{dt_0}{d\delta} = (f+t_0^2)^{\frac{1}{2}} J_f(-\delta t_0[f+t_0^2]^{-\frac{1}{2}}), \qquad \ldots\ldots(42)$$

where $$J_f(x) = \frac{Hh_{f-1}(x)}{fHh_f(x)}.$$

It may be shown that $$\frac{d}{dx}J_f(x) = fJ_f^2(x) - xJ_f(x) - 1. \qquad \ldots\ldots(43)$$

For given values of f and ϵ, the value of t_0 corresponding to $\delta = 0$ can be found in a straightforward manner since it is simply a probability level of Student's t. Starting from this initial value we can solve the differential equation (42) numerically and obtain the values of t_0 corresponding to non-zero values of δ. A trial of this method was made, taking $f = 6$, and $\epsilon = 0{\cdot}05$. The differential equation was solved by the Adams-Bashforth process. The values of higher derivatives of t with respect to δ at the point $\delta = 0$ are required in this process. By means of (43) the formulae for these quantities were obtained in quite a simple form.

This method gave promising results—it was suitable for large-scale work and there was comparatively little chance of inaccuracies entering the work—but it suffered from the defect that we could proceed to build up our table by small increments only of δ (actually increments of 0·05). As we wanted to be able to deal with large as well as small values of δ we decided that the process was not sufficiently speedy for our purposes.

It may be remarked, however, that a similar type of approach might be quite suitable in cases where the expression for the differential coefficient is simpler than was ours.

(iii) *Solution based on normal approximation*

The method finally adopted for calculating the greater part of the table is based on the use of Edgeworth's development of the normal probability function. In a recent paper E. A. Cornish & R. A. Fisher (1937) have given formulae, derived from this development, which make it possible to find percentage levels of a probability function when the cumulants of the function are given. These formulae are of most practical use when the probability function is itself nearly normal in form. With markedly non-normal distributions, the development may not converge, or at least more terms may be required than is practically convenient.

The cumulants of the non-central t-distribution are not difficult to find. The first three are given in equations (19), (20) and (21) and the values of $\sqrt{\beta_1}$ for different f and δ are given in Table I. It was pointed out in the discussion of Section 6 that the non-central t-distribution would often be markedly skew. It would appear, therefore, that it may not be a suitable distribution to express by a development of the normal function. Fortunately it is possible, as is pointed out in equation (26), to express problems demanding the quantities $t(f, \delta, \epsilon)$ and $\delta(f, t_0, \epsilon)$ in a form such that the solution can be based on a distribution much more nearly normal than that of non-central t.

It was shown that the statement

$$P(t > t_0 \mid f, \delta) = \epsilon$$

was equivalent to the statement

$$P(Y < \delta \mid f, t_0) = \epsilon,$$

where

$$Y = (-z) + t_0 \chi f^{-\frac{1}{2}}.$$

(Here z is a unit normal deviate and χ is distributed in the χ-distribution with f degrees of freedom.) The problem of finding $\delta(f, t_0, \epsilon)$ is therefore equivalent to finding a percentage level of Y.

The cumulants of Y may be obtained immediately from the formulae

$$\kappa_r(Y) = \kappa_r(-z) + f^{-\frac{1}{2}r} t_0^r \kappa_r(\chi). \qquad \text{......(44)}$$

In particular for $r = 2,\quad \kappa_2(Y) = 1 + f^{-1} t_0^2 \kappa_2(\chi),$

while for $r \neq 2,\quad \kappa_r(Y) = f^{-\frac{1}{2}r} t_0^r \kappa_r(\chi).$

Hence for all r the shape coefficients $\gamma_r(Y)$ increase from 0 to $\gamma_r(\chi)$ as t_0 increases from 0 to ∞. The γ's of Y are therefore always smaller than those of χ. But these latter are quite small even for small f. Hence the distribution of Y can never be far removed from normality. The distribution of Y is therefore suitable for development by Edgeworth's method and Cornish & Fisher's formulae may conveniently be used to provide percentage levels.

In this approach the problem of finding $\delta(f, t_0, \epsilon)$ appears as more fundamental than that of finding $t(f, \delta, \epsilon)$. Hence Table IV is designed to facilitate the direct calculation of $\delta(f, t_0, \epsilon)$. Its use to provide $t(f, \delta, \epsilon)$ demands a process of iteration which is not however difficult. The actual quantity $\lambda(f, t_0, \epsilon)$ which is tabled differs by a small amount from K_ϵ, the normal deviate which is exceeded with probability ϵ. This difference is given by the Cornish-Fisher formula. All the terms in this formula were used. This necessitated the use of the first six cumulants of χ. A method of obtaining these has already been described in a previous note (N. L. Johnson & B. L. Welch, 1939).

The greater part of the computation of Tables IV and V has been the work of one of us (N. L. Johnson). The remainder and much subsidiary checking work is due to Miss Catherine M. Thompson. We gratefully acknowledge her assistance.

REFERENCES

AIREY, J. R. (1931). "Table of *Hh* functions." *British Association Mathematical Tables.* Vol. I. London: Office of the British Association, Burlington House, W. 1.

CORNISH, E. A. & FISHER, R. A. (1937). *Rev. de l'Institut Intern. de Stat.* 5e Année, 307.

FISHER, R. A. (1931). "Properties of *Hh* functions." *Introduction to British Association Mathematical Tables*, Vol. I, xxvi, *loc. cit.*

JENNETT, W. J. & WELCH, B. L. (1939). *J. R. Statist. Soc. Suppl.* **6**, 80.

JOHNSON, N. L. & WELCH, B. L. (1939). *Biometrika*, **31**, 216.

McKAY, A. T. (1932). *J. R. Statist. Soc.* **95**, 695.

NEYMAN, J. (1935). *J. R. Statist. Soc. Suppl.* **2**, 131.

NEYMAN, J. & TOKARSKA, B. (1936). *J. Amer. Statist. Ass.* **31**, 318.

PEARSON, E. S. (1935). "The Application of Statistical Methods to Industrial Standardization and Quality Control." *B.S.S* 600. London: British Standards Institution.

MISCELLANEA

(i) Stirling's formula with remainder*

By E. V. HUNTINGTON, Harvard University

We assume the following theorem as known:

$$n\,! = \sqrt{(2\pi)}\sqrt{n}\,n^n\,e^{-n}[e^P],$$

where $n = 2, 3, 4, \ldots$, and

$$P = \frac{1}{12n} - \frac{1}{360n^3} + \frac{1}{1260n^5} - \frac{1}{1680n^7} + \frac{1}{1188n^9} - \frac{691}{360360n^{11}} + \ldots,$$

$$= 0{\cdot}083,333n^{-1} - 0{\cdot}002,777n^{-3} + 0{\cdot}000,794n^{-5} - 0{\cdot}000,595n^{-7}$$
$$+ 0{\cdot}000,842n^{-9} - 0{\cdot}001,918n^{-11} + \ldots.$$

We assume also that the error involved in stopping the series at any point is less (in absolute value) than the first term dropped. (For proof, see, for example, E. B. Wilson, *Advanced Calculus*, p. 456.)

Putting $Q = e^P$, and expanding and collecting terms, we find:

$$n\,! = \sqrt{(2\pi)}\sqrt{n}\,n^n\,e^{-n}[Q],$$

where

$$Q = 1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51,840n^3} - \frac{571}{2,488,320n^4}$$
$$+ \frac{163,879}{209,018,880n^5} + \frac{5,246,819}{75,246,796,800n^6} - \frac{534,703,531}{902,961,561,600n^7}$$
$$- \frac{4,483,131,259}{86,684,309,913,600n^8} + \frac{432,261,921,612,371}{514,904,800,886,784,000n^9} + \ldots,$$
$$= 1 + 0{\cdot}083,333,333n^{-1} + 0{\cdot}003,472,222n^{-2}$$
$$- 0{\cdot}002,681,327n^{-3} - 0{\cdot}000,229,472n^{-4}$$
$$+ 0{\cdot}000,784,039n^{-5} + 0{\cdot}000,069,728n^{-6} - 0{\cdot}000,592,166n^{-7}$$
$$- 0{\cdot}000,051,718n^{-8} + 0{\cdot}000,839,499n^{-9} + \ldots.$$

(For the terms up to and including n^{-7}, compare, for example, H. T. Davis, *Tables*, **1**, 180.)

Let Q_2 = the sum of the Q-series up to and including the term in n^{-2}, Q_3 = the sum of the Q-series up to and including the term in n^{-3}, etc.

Then we find:

$$Q = 1 + R_1, \quad \text{where} \quad 0{\cdot}083,33n^{-1} < R_1 < 0{\cdot}085,07n^{-1}.$$
$$Q = Q_1 + R_2, \quad \text{where} \quad 0{\cdot}002,07n^{-2} < R_2 < 0{\cdot}003,48n^{-2}.$$
$$Q = Q_2 + R_3, \quad \text{where} \quad -0{\cdot}002,80n^{-3} < R_3 < -0{\cdot}002,59n^{-3}.$$
$$Q = Q_3 + R_4, \quad \text{where} \quad -0{\cdot}000,23n^{-4} < R_4 < 0{\cdot}000,18n^{-4}.$$
$$Q = Q_4 + R_5, \quad \text{where} \quad 0{\cdot}000,66n^{-5} < R_5 < 0{\cdot}000,82n^{-5}.$$
$$Q = Q_5 + R_6, \quad \text{where} \quad -0{\cdot}000,24n^{-6} < R_6 < 0{\cdot}000,07n^{-6}.$$
$$Q = Q_6 + R_7, \quad \text{where} \quad -0{\cdot}000,62n^{-7} < R_7 < -0{\cdot}000,39n^{-7}.$$
$$Q = Q_7 + R_8, \quad \text{where} \quad -0{\cdot}000,06n^{-8} < R_8 < 0{\cdot}000,39n^{-8}.$$
$$Q = Q_8 + R_9, \quad \text{where} \quad -0{\cdot}000,01n^{-9} < R_9 < 0{\cdot}000,88n^{-9}.$$

* Presented to the American Mathematical Society, 8 April 1939.

(ii) A note on the interpretation of quasi-sufficiency

By M. S. BARTLETT

1. It has been shown by Fisher that if in problems of location we confine our attention to samples with the same configuration C, where C denotes the differences between the observations, then no statistical information in his sense is lost whatever reasonable statistic T we choose as an estimate of the unknown parameter θ. This result follows for probability laws of the type under consideration,

$$p(x \mid \theta) \equiv f(x-\theta)\,dx,$$

from the relation
$$p(S \mid \theta) = p(T \mid C, \theta)\,p(C), \qquad \text{......(1)}$$

where S denotes the sample of observations. For the conditional statistic $T \mid C$—a random or statistical variable T varying for given or fixed C—I have when a relation like (1) exists used the term *quasi-sufficient*, in view of the correspondence of (1) with the relation for a sufficient statistic T,

$$p(S \mid \theta) = p(T \mid \theta)\,p(S \mid T). \qquad \text{......(2)}$$

Conditional statistics also occur in the theory of more than one unknown, when the fixing of some statistical variables has the effect of eliminating unwanted parameters from our distributions. Confining our attention, however, to problems which are primarily problems in one unknown only, we require to examine relations of the type (1) further, in view of some recent comments by Welch* on the extent to which any conditional statistic like $T \mid C$ can claim to be sufficient.

From Fisher's theory of information, it follows that when estimating θ from two or more samples, we should naturally weight the different samples according to the information $I(C)$ in each. But Welch has pointed out that if we are concerned with interval estimation of θ from a single sample, or with the most efficient statistical test associated with a value θ_0, it does not appear that all our information, in the widest sense, is retained in $T \mid C$. If, for example, the only alternative to θ_0 happens to be θ_1, it is known that the appropriate criterion for discriminating θ_0 and θ_1 is $p(S \mid \theta_1)/p(S \mid \theta_0)$. While from equation (1) it follows that this criterion is equivalent to $p(T \mid C, \theta_1)/p(T \mid C, \theta_0)$, the former criterion must be referred to the distribution of S, not of $S \mid C$.

In partial answer to this criticism, it might be noticed that when considering $S \mid C$, we are not bound to choose the same significance level α for each C observed. Suppose we choose $\alpha(C_1) = \epsilon$ for the first sample. If the second sample gave a different configuration C_2, we might take $\alpha(C_2)$ such that the power of the test was a maximum for the two configurations C_1 and C_2 if we had chosen $\alpha(C_1)$ such that $\frac{1}{2}\{\alpha(C_1)+\alpha(C_2)\} = \epsilon$. Similarly for C_3, and so on. In the long run, the significance level adopted on this rule would be the average value $E_r\{\alpha_r(C_r)\}$, where $\alpha_r(C_r)$ denotes the significance level taken for the rth sample when making the rth test. For any finite number of samples the average significance level adopted is not exactly equal to ϵ, but the level becomes equal to ϵ in the limit; thus we could argue that, theoretically at least, we should eventually reach the most powerful test for a given significance level merely from consideration of $S \mid C$. In the above argument we allow the configuration distribution to be generated by the samples themselves. It is of course theoretical quibbling, for if we may assume that C has a definite distribution $p(C)$ which is already known, we should use this fact rather than wait for its confirmation by repeated sampling.

My own answer to the query of how far $T \mid C$ can be considered sufficient is that it is only sufficient provided that we do really agree to consider samples with the same configuration as that observed. The rejection altogether of the factor $p(C)$ has two consequences, (i) we have seen that a test might conceivably be derived on the basis of $p(C)$ more powerful than

* B. L. Welch. *Annals Math. Statist.* **10** (1939), 58–69.

the test based on $T \mid C$ with fixed significance level $\alpha(C) = \epsilon$, (iii) the fixing of C implies that whatever selection for C there might be distorting $p(C)$, the validity of tests based on $T \mid C$ will be unaffected.

The independence of the test from selection for C is the explicit advantage gained, though in addition it is possible that if our specification of the population is erroneous, that it is less so for $p(T \mid C)$ than $p(S)$; for example, if the configuration C is similar to the expected configuration on normal theory, it may prove more feasible to assume for a certain range of populations differing from normality that normal theory may still be approximately substituted for the unknown probability $p(T \mid C)$ than that we may do so for $p(S)$ or $p(T)$, though I have not investigated this point.

2. It is instructive to remember Fisher's comparison of the quasi-sufficient statistic $T \mid C$ with the sufficient statistic T, where T may be written in full, $T \mid n$, where n is the size of the sample. The equivalence of the two types of statistic when we regard C as fixed can be extended, in terms of our theoretical argument, if conversely we assume that n had in some problem a definite and known distribution from sample to sample. Consider similarly the test of significance of a regression coefficient. The orthodox theory is to consider the conditional statistic $b \mid \Sigma(x-\bar{x})^2$, where b is our estimate, and $\Sigma(x-\bar{x})^2$ the sum of squares of deviations of the independent variable x. Suppose for the sake of argument that the true variance of the residual dependent variable y_x was known to be unity, and the x's are such that $\Sigma(x-\bar{x})^2 = 1$ on Mondays and 1·44 on Tuesdays. Then for an 0·025 significance level (one tail), the usual practice would be to take 1·96 as the significance level for b (from $b_0 = 0$) on Mondays, and $1{\cdot}96/1{\cdot}2 = 1{\cdot}633$ on Tuesdays. The power of the test in relation to the alternative that $b_1 = 3{\cdot}92$ is 0·9860. But if we were satisfied with adjusting the significance level to be 0·025 merely in the long run for Mondays and Tuesdays together, we may raise the power of the test to its maximum value of 0·9878, by taking the Monday significance level at $b = 1{\cdot}87$ ($\alpha = 0{\cdot}0307$) and the Tuesday level at $b = 1{\cdot}723$ ($\alpha = 0{\cdot}0194$).

Summary

In a theoretical note on the interpretation of quasi-sufficient statistics in problems of one unknown parameter, it is agreed with Welch that full sufficiency properties can only be claimed if we deliberately confine our attention to conditional samples of the type observed; but it is pointed out that the resulting inferences are independent of any selection operating on the variables that have been fixed.

(iii) The cumulants and moments of the binomial distribution, and the cumulants of χ^2 for a $(n \times 2)$-fold table

By J. B. S. HALDANE, F.R.S.

From the Department of Biometry, University College, London

The first four cumulants of the distribution of χ^2 for a $(n \times 2)$-fold table when samples are finite, have already been given (Haldane, 1937). These and higher cumulants and moments can be calculated by a simpler method. Consider a sample of s, the probability of a success being p, and $p+q = 1$. Pearson (1919) pointed out that for moments of $(p+q)^s$ about its mean, the generating function is $(qe^{pt}+pe^{-qt})^s$, and Romanovsky (1923) gave a recurrence formula for the moments. That for the cumulants is much simpler.

Let $$U = qe^{pt}+pe^{-qt}.$$

Then the cumulant-generating function

$$K(t) \equiv \sum_{r=2}^{\infty} \frac{\kappa_r t^r}{r!} = s \log U.$$

To find the cumulants for $s = 1$, we note that

$$\frac{\partial}{\partial q} K(t) = \frac{e^{pt} - e^{-qt}}{U} - t,$$

$$\frac{\partial}{\partial t} K(t) = \frac{pq(e^{pt} - e^{-qt})}{U}$$

So

$$\frac{\partial}{\partial t} K(t) = pq \frac{\partial}{\partial q} K(t) + pqt,$$

or

$$\sum_{r=2}^{\infty} \frac{\kappa_r t^{r-1}}{(r-1)!} \equiv pq \sum_{r=2}^{\infty} \frac{t^r}{r!} \frac{d\kappa_r}{dq} + pqt.$$

Equating the coefficients of $t^r/r!$, we find

$$\kappa_2 = pq,$$

and if $r > 2$

$$\kappa_{r+1} = pq \frac{d\kappa_r}{dq}. \qquad \text{......(1)}$$

Let $pq = c$, $p - q = g$; then

$$\kappa_{r+1} = c \frac{d\kappa_r}{dq}, \quad \frac{dc}{dq} = g, \quad \frac{dg}{dq} = -2, \quad g^2 = 1 - 4c.$$

Hence if $\kappa_{2r} = f(c)$,

$$\left.\begin{aligned} \kappa_{2r+1} &= gcf'(c), \\ \kappa_{2r+2} &= c(1-6c)f'(c) + c^2(1-4c)f''(c). \end{aligned}\right\} \qquad \text{......(2)}$$

From these equations we can very rapidly calculate successive values of κ_r, since $\kappa_2 = c$, and find:

$$\begin{aligned}
\kappa_1 &= 0, \\
\kappa_2 &= c, \\
\kappa_3 &= cg, \\
\kappa_4 &= c - 6c^2, \\
\kappa_5 &= g(c - 12c^2), \\
\kappa_6 &= c - 30c^2 + 120c^3, \\
\kappa_7 &= g(c - 60c^2 + 360c^3), \\
\kappa_8 &= c - 126c^2 + 1{,}680c^3 - 5{,}040c^4, \\
\kappa_9 &= g(c - 252c^2 + 5{,}040c^3 - 20{,}160c^4), \\
\kappa_{10} &= c - 510c^2 + 17{,}640c^3 - 151{,}200c^4 + 362{,}880c^5, \\
\kappa_{11} &= g(c - 1{,}020c^2 + 52{,}920c^3 - 604{,}800c^4 + 1{,}814{,}400c^5), \\
\kappa_{12} &= c - 2{,}046c^2 + 168{,}960c^3 - 3{,}160{,}080c^4 + 19{,}958{,}400c^5 - 39{,}916{,}800c^6. \qquad \text{......(3)}
\end{aligned}$$

If each of the above cumulants be multiplied by s, the moments about the mean can now be calculated from the expressions given by Fisher (1928) and Haldane (1938). If $p = \frac{1}{2}$ we have

$$K(t) = s \log \cosh \tfrac{1}{2}t,$$

so

$$\kappa_2 = \frac{s}{4}, \quad \kappa_4 = -\frac{s}{8}, \quad \kappa_6 = \frac{s}{4}, \quad \kappa_8 = -\frac{17s}{16}, \quad \kappa_{10} = \frac{31s}{4}, \quad \kappa_{12} = -\frac{691s}{8},$$

while if q is very small we have for the cumulant-generating function of a Poisson series

$$K(t) = sc(e^t - 1 - t).$$

The coefficient of c^2 is $-[s^r+(-1)^r-3]$. So when q is small, but its square is not neglected, the first order correction to the Poisson cumulant-generating function is

$$K(t) = sq(e^t-1-t)+sq^2(e^{2t}-e^t).$$

The numerical coefficient of the highest power of c in κ_r is $(r-1)!$ when r is even, and $\frac{1}{2}(r-1)!$ when r is odd.

Consider a sample of s, in which a successes are recorded. Then

$$\chi^2 = \frac{(a-sp)^2}{spq}.$$

But $a-sp$ is the departure from the mean of the binomial distribution $(p+q)^s$. Hence the rth moment of the distribution of χ^2 (for one degree of freedom) about zero, is

$$\nu'_r = \frac{\mu_{2r}}{s^r c^r},$$

where μ_{2r} is the $2r$th moment of $(p+q)^s$.

But if μ'_r and κ'_r be the rth moment about the mean, and the rth cumulant, of the χ^2 distribution, then

$$\mu_2 = \nu'_2-\nu'^2_1, \text{ etc.,} \quad \kappa'_2 = \mu'_2, \text{ etc.}$$

Making the necessary substitutions, we find, for the cumulants of χ^2 in terms of those of the binomial distribution:

$$\begin{aligned}
\kappa'_1 &= (sc)^{-1}\kappa_2,\\
\kappa'_2 &= (sc)^{-2}(2\kappa_2^2+\kappa_4),\\
\kappa'_3 &= (sc)^{-3}[8\kappa_2^3+2(5\kappa_3^2+6\kappa_2\kappa_4)+\kappa_6],\\
\kappa'_4 &= (sc)^{-4}[48\kappa_2^4+48(5\kappa_2\kappa_3^2+3\kappa_2^2\kappa_4)+8(4\kappa_4^2+7\kappa_3\kappa_5+3\kappa_2\kappa_6)+\kappa_8],\\
\kappa'_5 &= (sc)^{-5}[384\kappa_2^5+960(5\kappa_2^2\kappa_3^2+2\kappa_2^3\kappa_4)+80(25\kappa_3^2\kappa_4+16\kappa_2\kappa_4^2+28\kappa_2\kappa_3\kappa_5\\
&\qquad +6\kappa_2^2\kappa_6)+2(63\kappa_5^2+100\kappa_4\kappa_6+60\kappa_3\kappa_7+20\kappa_2\kappa_8)+\kappa_{10}],\\
\kappa'_6 &= (sc)^{-6}[3{,}840\kappa_2^6+9{,}600(10\kappa_2^3\kappa_3^2+3\kappa_2^4\kappa_4)+4{,}800(3\kappa_3^4+25\kappa_2\kappa_3^2\kappa_4\\
&\qquad +8\kappa_2^2\kappa_4^2+14\kappa_2^2\kappa_3\kappa_5+2\kappa_2^3\kappa_6)+40(132\kappa_4^3+672\kappa_3\kappa_4\kappa_5+189\kappa_2\kappa_5^2\\
&\qquad +226\kappa_3^2\kappa_6+300\kappa_2\kappa_4\kappa_6+180\kappa_2\kappa_3\kappa_7+30\kappa_2^2\kappa_8)+4(113\kappa_6^2+198\kappa_5\kappa_7\\
&\qquad +120\kappa_4\kappa_8+55\kappa_3\kappa_9+15\kappa_2\kappa_{10})+\kappa_{12}]. \qquad \ldots\ldots(4).
\end{aligned}$$

We now substitute the values of κ_r given in equations (3) multiplied by s, putting

$$k = (pq)^{-1} = c^{-1}.$$

We therefore have, for the cumulants of χ^2 *with one degree of freedom*:

$$\begin{aligned}
\kappa_1 &= 1,\\
\kappa_2 &= 2+(k-6)s^{-1},\\
\kappa_3 &= 8+2(11k-56)s^{-1}+(k^2-30k+120)s^{-2},\\
\kappa_4 &= 48+96(4k-19)s^{-1}+16(7k^2-125k+420)s^{-2}+(k^3-125k^2+1{,}680k-5{,}040)s^{-3},\\
\kappa_5 &= 384+960(7k-32)s^{-1}+400(15k^2-214k+648)s^{-2}+6(81k^3\\
&\qquad -3{,}908k^2+38{,}420k-98{,}496)s^{-3}+(k^4-510k^3+17{,}640k^2\\
&\qquad\qquad -151{,}200k+362{,}880)s^{-4},\\
\kappa_6 &= 3{,}840+9{,}600(13k-58)s^{-1}+9{,}600(26k^2-327k+924)s^{-2}\\
&\qquad +40(1{,}729k^3-56{,}236k^2+459{,}024k-1{,}065{,}792)s^{-3}+4(501k^4\\
&\qquad -59{,}398k^3+1{,}289{,}244k^2-8{,}824{,}320k+18{,}555{,}840)s^{-4}\\
&\qquad +(k^5-2{,}046k^4+168{,}960k^3-3{,}160{,}080k^2+19{,}958{,}400k\\
&\qquad\qquad -39{,}916{,}800)s^{-5}. \qquad \ldots\ldots(5)
\end{aligned}$$

When $p = \frac{1}{2}$, $k = 4$, and we have, for n degrees of freedom:

$$\begin{aligned}
\kappa_1 &= n,\\
\kappa_2 &= 2ns^{-1}(s-1),\\
\kappa_3 &= 8ns^{-2}(s-1)(s-2),\\
\kappa_4 &= 16ns^{-3}(s-1)(3s^2-15s+17),\\
\kappa_5 &= 128ns^{-4}(s-1)(s-2)(3s^2-21s+31).\\
\kappa_6 &= 256ns^{-5}(s-1)(15s^4-210s^3+990s^2-1{,}950s+1{,}382). \qquad \ldots\ldots(6)
\end{aligned}$$

If there are n samples, with different values of s, we have, for the cumulants of χ^2, where $h = \dfrac{1}{2pq}$, and $R_i = \Sigma s^{-i}$,

$$\begin{aligned}
\kappa_1 &= n,\\
\kappa_2 &= 2n[1+(h-3)R_1],\\
\kappa_3 &= 4n[2+(11h-28)R_1+(h^2-15h+30)R_2],\\
\kappa_4 &= 8n[6+12(8h-19)R_1+4(14h^2-125h+210)R_2+(h^3-63h^2+420h-630)R_3],\\
\kappa_5 &= 16n[24+120(7h-16)R_1+100(15h^2-107h+162)R_2+3(81h^3\\
&\quad -1{,}954h^2+9{,}560h-12{,}312)R_3+(h^4-255h^3+4{,}410h^2-18{,}900h+22{,}680)R_4],\\
&= 32n[120+600(13h-29)R_1+600(52h^2-327h+462)R_2+10(1{,}729h^3\\
&\quad -28{,}118h^2+114{,}756h-133{,}228)R_3+2(501h^4-29{,}699h^3+322{,}311h^2\\
&\quad -1{,}103{,}040h+1{,}159{,}740)R_4+(h^5-1{,}023h^4+42{,}240h^3-395{,}010h^2\\
&\quad +1{,}247{,}400h-1{,}247{,}400)R_5]. \qquad \ldots\ldots(7)
\end{aligned}$$

When $p = q = \frac{1}{2}$, we have:

$$\begin{aligned}
\kappa_1 &= n,\\
\kappa_2 &= 2(n-R_1),\\
\kappa_3 &= 8(n-3R_1+4R_2),\\
\kappa_4 &= 16(3n-18R_1+32R_2-17R_3),\\
\kappa_5 &= 128(3n-30R_1+100R_2-135R_3+62R_4),\\
\kappa_6 &= 256(15n-225R_1+1{,}200R_2-2{,}940R_3+3{,}332R_4-1{,}382R_5). \qquad \ldots\ldots(8)
\end{aligned}$$

The first four of equations (5, 6, 7, 8) have already been given in a slightly different form by Haldane (1937). The limiting forms of equations (5) and (7) when s tends to infinity and k to zero, while $ks = g$, have been given by Haldane (1938). However, the expression for κ_6 there given is incorrect. The coefficient of R_1 in the expression for κ_6 should be 124,800.

The extension of equations (7) would be rather tedious. However, those of equations (6) and (8) would not be very difficult. The coefficient of $x^{2r}/2r!$ in the expansion of $\log\cosh t$ is the value of $(d/dx)^{2r-1}(1-\tanh^2 x)$ when $x = 0$, and can easily be calculated, since this differential coefficient is a polynomial in $\tanh x$. The equations for moments in terms of cumulants can easily be extended when all odd cumulants vanish. In this case a useful check can be obtained from the fact that when $s = 2$ the cumulant-generating function of χ^2 is $t+\log\cosh t$.

Summary

Expressions are obtained for the first twelve cumulants of the binomial distribution, and a simple recurrence formula for further cumulants. The first six cumulants of χ^2 for a $(n \times 2)$-fold table when expectations are small, are deduced.

REFERENCES

FISHER, R. A. (1928). "Moments and product moments of sampling distributions." *Proc. Lond. Math. Soc.* **30**, 200–38.

HALDANE, J. B. S. (1937). "The exact value of the moments of the distribution of χ^2, used as a test of goodness of fit, when expectations are small." *Biometrika*, **29**, 133–43.

—— (1938). "The first six moments of χ^2 for an n-fold table with n degrees of freedom when some expectations are small." *Biometrika*, **29**, 389–91.

PEARSON, K. (1919). "Peccavimus" (Footnote, p. 270). *Biometrika*, **12**, 259–81.

ROMANOVSKY, V. (1923). "Note on the moments of a binomial $(p+q)^s$ about its mean." *Biometrika*, **15**, 410–12.

CAUSATION AND CORRELATION

(iv) **The principles of the mathematical theory of correlation.** By A. A. TSCHUPROW. Translated by M. KANTOROWITSCH. Wm. Hodge & Co. Ltd. Price 12*s*. 6*d*.

This book is a translation of an enlarged reproduction of lectures delivered by Professor Tschuprow at the University of Oslo, and originally printed in German about a decade ago. It will be of interest to many people in that it is a complete survey of correlation theory and its underlying principles. The theory is expounded along the now familiar classical lines which had their origin in Karl Pearson's writings, and although the application of this theory to practical problems is perhaps a little out of date, it still forms a necessary background which the student must acquire, and of which a proper understanding will always be essential.

The book as a whole shows an astonishing "patchiness" in writing. It may be the fault of the translator, but the fact remains that the meaning of whole paragraphs is sometimes very obscure, while at other times the ease and lucidity with which arguments are presented are unrivalled by any comparable treatise. This unevenness is unfortunate, since the book will be read more for the exposition of underlying principles than for its algebraic development of the theory; indeed it may be questioned whether those who have not previously mastered the elements of statistical theory and probability will obtain any profit from its reading. There is no clear discussion of probability and what it means, and the reader is left to find out from examples how a probability may be calculated. This cannot be deemed a fault, but seems to point to the book being useful only if previous knowledge of the subject is obtained elsewhere.

The chapter on stochastic connexion and functional relationship is good, and may be read with advantage by any statistical worker and teacher. The method of approach is the same as that of the present day, and will doubtless be followed for many years to come. It is, however, a little astonishing to find that the translator makes no use of the term "random variable", which has long passed into common use. The general discussion throughout the book and in particular in the chapter entitled "Object and value of Correlation Measurement" is stimulating, even if the reader is not always in agreement with the author. The mathematical development of the theory is set out with clarity and freshness, which make it enjoyable reading. The notation is a little cumbersome, but, once mastered, it does not prove difficult to follow.

Taken as a whole, this book is a worthy contribution to correlation literature, and it is surprising that no translation has been published until this date. It certainly should be read by all who attempt to gain an understanding of statistical theory.

F. N. DAVID.